

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:03 : Search time 43.6471 Seconds
(without alignments)
285.020 Million cell updates/sec

Title: US-10-032-658-11
Perfect score: 664
Sequence: 1 MARFCTGFSKKMLVIAVIWM.....DSNCKYKATACNTSGCPGH 112

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	664	100.0	112	20	AAV09278
2	603	90.8	124	20	AAV09281
3	599	90.2	112	20	AAV09279
4	599	90.2	124	20	AAV09282
5	487	73.3	148	20	AAV09280
6	377.5	56.9	108	18	AAW07693
7	361	54.4	109	18	AAW07694
8	156	23.5	1679	22	AAU07343
9	139	20.9	233	21	AAV74791
10	124	18.7	24	20	AAV09276
11	123.5	18.6	524	22	AAU07370

12	119	17.9	250	21	AAV82335
13	118.5	17.8	2189	11	AAV05222
14	116	17.5	1700	21	AAV18144
15	111	16.7	225	21	AAV36084
16	110.5	16.6	233	20	AAV29053
17	110.5	16.6	233	22	AAU25524
18	110.5	16.6	384	17	AAV05479
19	109.5	16.5	1193	12	AAW05835
20	109.5	16.5	1193	21	AAV59599
21	109	16.4	124	21	AAV82334
22	107.5	16.2	2703	22	ABB60266
23	107	16.1	250	22	AAV79002
24	107	16.1	301	22	AAV79986
25	107	16.1	1712	13	AAV22461
26	106	16.0	229	22	AAV79404
27	105.5	15.9	1055	19	AAW44298
28	105.5	15.9	1148	20	AAV87895
29	105.5	15.9	1212	19	AAW44299
30	105.5	15.9	1257	17	AAW05834
31	105.5	15.9	1257	21	AAV59598
32	105	15.8	727	11	AAV05533
33	105	15.8	762	21	AAV92455
34	105	15.8	3680	22	ABB70878
35	104	15.7	154	22	AAV78420
36	104	15.7	167	22	AAV39466
37	104	15.7	222	22	AAW41252
38	103.5	15.6	787	18	AAW13573
39	103	15.5	718	11	AAV05936
40	103	15.5	762	21	AAV92448
41	103	15.5	762	21	AAV92449
42	103	15.5	762	21	AAV92450
43	103	15.5	762	21	AAV92451
44	103	15.5	762	21	AAV92452
45	103	15.5	762	21	AAV92453

ALIGNMENTS

RESULT 1	AAV09278	standard; Protein; 112 AA.
ID	AAV09278:	
AC	AAV09278:	
XX		
DT	06-JUL-1999	(first entry)
XX		
DE	YL-1 thermal hysteresis protein.	
XX		
KW	Antifreeze; thermal hysteresis protein; THP; mealworm beetle; Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.	
XX		
OS	Tenebrio molitor.	
XX		
PN	WO9900493-A1.	
XX		
PD	07-JAN-1999.	
XX		
PF	25-JUN-1998:	98WO-CA00618.
XX		
PR	26-JUN-1997:	97US-0882907.
XX		
PA	(TMOH) UNIV QUEBENS KINGSTON.	
XX		
PI	Davies PL, Graham LA, Liou Y, Walker VK;	
XX		
DR	WPI, 1999-095739/08.	
XX	N-PSDB: AAX33465.	
XX		
PT	New nucleic acid encoding antifreeze proteins of Tenebrio molitor -	
XX	used to improve low temperature tolerance of fish, plants etc., and	
XX	quality of foods or biological materials during frozen storage	

Applicant's
own work

PS Disclosure: Page 60; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD; (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c) (1) binds specifically to antibodies raised against the Tenebrio molitor YLI-Y14 antifreeze proteins, (11) is at least 60% identical with YLI-4 or (111) has a sequence that includes at least one repeating unit of the consensus sequence CTX₃XXCX₃AX₃ (I). (II), optionally present in an organism, are used to improve preservation of frozen foods (specifically to reduce formation of ice crystals following exposure to repeated freeze-thaw cycles) or viability of biological materials (e.g. organs, cells or extracts) stored or transported at low temperature, or more generally to depress the freezing point of any aqueous solution, particularly where this is applied to an organism. (I) is used to produce transgenic animals and plants (e.g. grapes, oilseed crops, grains, citrus fruit or sugar cane) having better tolerance to freezing. CC Fragments of (I) are used as probes to detect or isolate coding sequences, in diagnosis (for determining gene expression) and as CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to CC detect (II) is usual immunoassays. (II) from T. molitor have specific CC activity 100 times greater than that of fish antifreeze proteins. The CC present sequence represents a thermal hysteresis protein (THP) from the CC present invention.

XX Sequence 112 AA;

SQ Query Match 100.0%; Score 664; DB 20; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.6e-45;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFTGSGFKMWLIYAVIMCLCTECYCHCTGADCTGCTDAGTGCNCNPAHCTDTSKN 60
DB 1 MAFTGSGFKMWLIYAVIMCLCTECYCHCTGADCTGCTDAGTGCNCNPAHCTDTSKN 60
DB 61 CVKAACTGCTGKNTARTCTNSKDCFEAKTCTDSTNCYKATACNPGCGH 112
DB 61 CVKAACTGCTGKNTARTCTNSKDCFEAKTCTDSTNCYKATACNPGCGH 112

RESULT 2
ID AAY09281 standard; Protein; 124 AA.

XX AAY09281;
XX 06-JUL-1999 (first entry)
XX YL-3 thermal hysteresis protein.
XX Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
XX Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX Tenebrio molitor.
XX OS
XX PN W09900493-A1.
XX PD 07-JAN-1999.
XX PF 25-JUN-1998; 98WO-CA00618.
XX PR 26-JUN-1997; 97US-0882907.
XX (TOOH) UNIV QUEENS KINGSTON.
XX PA Davies PL, Graham LA, Liou Y, Walker VK;
XX PI WPI: 1999-095739/08.
XX DR N-PSDB; AAX33468.
XX PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
XX PT used to improve low temperature tolerance of fish, plants etc., and
XX PT quality of foods or biological materials during frozen storage

XX Disclosure: Page 66; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD; (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c) (1) binds specifically to antibodies raised against the Tenebrio molitor YLI-Y14 antifreeze proteins, (11) is at least 60% identical with YLI-4 or (111) has a sequence that includes at least one repeating unit of the consensus sequence CTX₃XXCX₃AX₃ (I). (II), optionally present in an organism, are used to improve preservation of frozen foods (specifically to reduce formation of ice crystals following exposure to repeated freeze-thaw cycles) or viability of biological materials (e.g. organs, cells or extracts) stored or transported at low temperature, or more generally to depress the freezing point of any aqueous solution, particularly where this is applied to an organism. (I) is used to CC produce transgenic animals and plants (e.g. grapes, oilseed crops, CC grains, citrus fruit or sugar cane) having better tolerance to freezing. CC Fragments of (I) are used as probes to detect or isolate coding CC sequences, in diagnosis (for determining gene expression) and as CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to CC detect (II) is usual immunoassays. (II) from T. molitor have specific CC activity 100 times greater than that of fish antifreeze proteins. The CC present sequence represents a thermal hysteresis protein (THP) from the CC present invention.

XX Sequence 124 AA;

SQ Query Match 90.8%; Score 603; DB 20; Length 124;
Best Local Similarity 83.9%; Pred. No. 1.1e-40;
Matches 104; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 MAFTGSGFKMWLIYAVIMCLCTECYCHCTGADCTGCTDAGTGCNCNPAHCTDTSKN 59
DB 1 MAFTGSGFKMWLIYAVIMCLCTECYCHCTGADCTGCTDAGTGCNCNPAHCTDTSKN 60
QY 60 -----NCVAACTGCTGKNTARTCTNSKDCFEAKTCTDSTNCYKATACNPGCGH 108
DB 61 CVKAACTGCTGKNTARTCTNSKDCFEAKTCTDSTNCYKATACNPGCGH 120
QY 109 CPGH 112
DB 121 CPGH 124

RESULT 3
ID AAY09279 standard; Protein; 112 AA.

XX AAY09279;
XX 06-JUL-1999 (first entry)
XX YL-2 thermal hysteresis protein.
XX Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
XX Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX Tenebrio molitor.
XX OS
XX PN W09900493-A1.
XX PD 07-JAN-1999.
XX PF 25-JUN-1998; 98WO-CA00618.
XX PR 26-JUN-1997; 97US-0882907.
XX (TOOH) UNIV QUEENS KINGSTON.
XX PA Davies PL, Graham LA, Liou Y, Walker VK;
XX PI WPI: 1999-095739/08.
XX DR

XX (TOOH) UNIV QUEBENS KINGSTON.
 PA
 XX
 PI Davies PL, Graham LA, Liou Y, Walker VK.
 XX
 DR WPI: 1999-095739/08.
 XX N-PSDB: AAX33467.
 DR
 XX
 PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
 PT used to improve low temperature tolerance of fish, plants etc., and
 PT quality of foods or biological materials during frozen storage
 XX
 PS Disclosure; Page 64; 88pp; English.
 XX
 CC The present invention describes a nucleic acid (I) encoding an antifreeze
 CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
 CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)
 CC (i) binds specifically to antibodies raised against the Tenebrio molitor
 CC YLI-Y14 antifreeze proteins, (ii) is at least 60% identical with YLI-4
 CC or (iii) has a sequence that includes at least one repeating unit of the
 CC consensus sequence CRYSXKXAXT (I). (II), optionally present in an
 CC organism, are used to improve preservation of frozen foods (specifically
 CC to reduce formation of ice crystals following exposure to repeated
 CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
 CC cells or extracts) stored or transported at low temperature, or more
 CC generally to depress the freezing point of any aqueous solution,
 CC particularly where this is applied to an organism. (I) is used to
 CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
 CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
 CC Sequences of (I) are used as probes to detect or isolate coding
 CC sequences, in diagnosis (for determining gene expression) and as
 CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
 CC detect (II) is usual immunoassays. (II) from T. molitor have specific
 CC activity 100 times greater than that of fish antifreeze proteins. The
 CC present sequence represents a thermal hysteresis protein (THP) from the
 CC present invention.
 CC
 XX
 XX Sequence 148 AA:
 SO
 Query Match 73.3%; Score 487; DB 20; Length 148;
 Best Local Similarity 60.8%; Pred. No. 1.8e-31;
 Matches 90; Conservative 5; Mismatches 17; Indels 36; Gaps 2;
 OY 1 MARKTGFSKMLVIVWICLTGCTGAGDCTGCTGAGDCTGCGNCPNNAHTCTGSKN 60
 DB 1 MSKRISFTFKMLIVWICLTGCTGAGDCTGCTGAGDCTGCGNCPNNAHTCTGSKN 60
 OY 61 CVKAACTGCTGSKN-----TARTCTNSKDCFE----- 87
 DB 61 CVRATCTGCTGSKNRAATCTGCTGAGDCTGCTGAGDCTGCGNCPNNAHTCTGSKN 120
 OY 88 ---AKCTDSTNCKYKATCTNSTGCGPH 112
 DB 121 CYTATCTNSTNCKYKATCTNSTGCGPH 148
 RESULT 6
 AAM07693
 ID AAM07693 standard; Protein; 108 AA.
 AC AAM07693;
 XX
 DT 09-SEP-1997 (first entry)
 XX
 DE Dendroides canadensis thermal-hysteresis protein.
 XX
 KW THP; antifreeze; transformation; plant.
 XX
 OS Dendroides canadensis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /label= signal

FT Protein 20..108
 FT /product= Thermal-hysteresis_protein
 FT Region 20..32
 FT /label= Repeat_A
 FT /note= "See AAM07696 for 13 amino acid consensus
 FT sequence" Region 33..45
 FT /label= Repeat_B
 FT /note= "See AAM07696 for 13 amino acid consensus
 FT sequence" Region 46..58
 FT /label= Repeat_C
 FT /note= "See AAM07696 for 13 amino acid consensus
 FT sequence" Region 59..70
 FT /label= Repeat_D
 FT /note= "See AAM07697 for 12 amino acid consensus
 FT sequence" Region 7182..82
 FT /label= Repeat_E
 FT /note= "See AAM07697 for 12 amino acid consensus
 FT sequence" Region 83..94
 FT /label= Repeat_F
 FT /note= "See AAM07697 for 12 amino acid consensus
 FT sequence" Region 95..104
 FT /label= Repeat_G
 FT /note= "See AAM07697 for 12 amino acid consensus
 FT sequence"
 FT
 FT W09640973-A1.
 FT
 FT 19-DEC-1996.
 FT
 FT 05-JUN-1996; 96WO-0508815.
 FT
 FT 07-JUN-1995; 95US-0485359.
 FT 08-DEC-1995; 95US-0565954.
 FT
 FT (UNOT) UNIV NOTRE DAME DU LAC.
 FT
 FT Duman JG;
 FT
 FT WPI: 1997-052352/05.
 FT N-PSDB: AAT47150.
 FT
 FT DNA encoding Dendroides sp. thermal hysteresis protein - produces
 FT protein having antifreeze properties, useful to protect plant cells,
 FT seeds or plants from frost damage
 FT
 FT Disclosure; Page 31-32; 46pp; English.
 FT
 FT The present sequence represents a novel peptide which has antifreeze
 FT properties. The nucleic acid sequence was derived from Dendroides
 FT canadensis, and this protein is an example of a thermal-hysteresis
 FT protein. The mature protein sequence contains a 13 amino acid sequence
 FT and a 12 amino acid sequence that are repeated three and four times
 FT respectively (see features table). The nucleic acid sequence can
 FT be used to transform a plant cell for the heterologous production of
 FT Dendroides thermal hysteresis protein (THP), i.e. to protect it from
 FT frost damage. The DNA and polynucleotide sequences may be used as probes
 FT for the isolation of THP coding sequences, from other organisms.
 FT
 SO Sequence 108 AA;
 Query Match 56.9%; Score 377.5; DB 18; Length 108;
 Best Local Similarity 64.6%; Pred. No. 5.9e-23;
 Matches 64; Conservative 13; Mismatches 21; Indels 1; Gaps 1;
 OY 13 LVIAVWICLTGCTGAGDCTGCTGAGDCTGCGNCPNNAHT-CTDSKNCVKAATCTGCT 71
 DB 10 LVISVWLVWICHTGCTGAGDCTGCTGAGDCTGCTGAGDCTGCGNCPNNAHTCTGCT 69
 OY 72 KONTARTCTNSKDCFEAKTCTDSTNCKYKATCTNSTGCP 110
 DB 70 DCHNAETCTNSTNCKYKATCTGCTGAGDCTGCTGAGDCTGCGNCPNNAHTCTGCT 108

RESULT 7
ID AAM07694
AAW07694 standard; Protein; 109 AA.
XX AC AAW07694;
DT 09-SEP-1997 (first entry)
DE Dendroides canadensis thermal-hysteresis protein.
XX THP; antifreeze; transformation; plant.
OS Dendroides canadensis.
FH Key Location/Qualifiers
FT Peptide 1..19
FT /label= Signal
FT Protein 20..108
FT /product= Thermal-hysteresis-protein
FT Region 20..32
FT /label= Repeat_A
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 33..45
FT /label= Repeat_B
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 46..58
FT /label= Repeat_C
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 59..70
FT /label= Repeat_D
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 7182..82
FT /label= Repeat_E
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 83..94
FT /label= Repeat_F
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 95..104
FT /label= Repeat_G
FT /note= "See AAW07697 for 12 amino acid consensus
sequence"
PN W09640973-A1.
PD 19-DEC-1996.
XX 05-JUN-1996; 96WO-US08815.
PR 07-JUN-1995; 95US-0485359.
PR 08-DEC-1995; 95US-0569594.
XX (UNOT) UNIV NOTRE DAME DU LAC.
XX Duman JG;
P1 MPI; 1997-052352/05.
DR N-PSDB; AAT47151.
XX DNA encoding Dendroides sp. thermal hysteresis protein - produces
PT protein having antifreeze properties, useful to protect plant cells,
PS seeds or plants from frost damage
PS Disclosure; Page 33; 46pp; English.
XX The present sequence represents a novel peptide which has antifreeze
CC properties. The nucleic acid sequence was derived from Dendroides
CC canadensis, and this protein is an example of a thermal-hysteresis
CC protein. The mature protein sequence contains a 13 amino acid sequence
CC and a 12 amino acid sequence that are repeated three and four times
CC respectively (see features table). The nucleic acid sequence can
CC be used to transform a plant cell for the heterologous production of
CC dendroides thermal hysteresis protein (THP), i.e. to protect it from
CC frost damage. The DNA and polynucleotide sequences may be used as probes

[illegible]

Best Local Similarity 36.3%; Pred. No. 0.00019;
Matches 37; Conservative 1; Mismatches 44; Indels 20; Gaps 6;

QY 23 CTECYCHCTGADCTSCDADCTGC-----GNCNPAHCTDCKNVKAAAT-----CTGSTR 72
Db 1118 CTTCCCTGTGGCGCTACG-ACGGCCCTGCTGTGGTCCAGTACAGATGGACAGCTG 1176
QY 73 CNTARTCTNSKDCFEAKTCTDSTNCYKATACCTNSTG--CPG 111
Db 1177 C-----TCCGGACCGGTGACTGATC--AACCAAGTGTACTCG 1212

RESULT 9
AAV74791
ID AAV74791 standard; Protein: 233 AA.
AC AAV74791;
XX
XX
DT 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 263 protein sequence SEQ ID NO:1056.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
OS
XX MO9957280-A2.
XX
XX 11-NOV-1999.
PD
XX 30-APR-1999; 99WO-US09346.
PF
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098894.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR N-PSDB: AA253553.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX Claim 2; Page 606; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 233 AA;

Query Match 20.9%; Score 139; DB 21; Length 233;
Best Local Similarity 38.7%; Pred. No. 0.0074;
Matches 36; Conservative 3; Mismatches 42; Indels 12; Gaps 7;

QY 21 CLCTECYCHCTG-GADCTSCDADCTGCGNCPNPAHCTDCKNVKAAATG-TGST--KCMYA 76
Db 101 CATTTCGATGGCGGCTACCAACCAAGCAGC--AGGCAGI--CGAAGTGTGATGGCCGTA 156
QY 77 RCTCTNSKDCFEAKTCTDSTNCYKATACCTNSTGC 109
Db 157 SYCTGG--C--AACCCCTGTGCAACTACGTCAAC 185

RESULT 10
AAV09276
ID AAV09276 standard; peptide: 24 AA.
AC AAV09276;
XX
XX 06-JUL-1999 (first entry)
DT
XX Antifreeze protein amino acid sequence SEQ ID NO:4.
DE
XX Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
XX Synthetic.
OS
XX Tenebrio molitor.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..24
FT
FT
XX
XX WO9900493-A1.
XX
XX 07-JAN-1999.
PD
XX 25-JUN-1998; 98WO-CA00618.
PF
XX 26-JUN-1997; 97US-0882907.
PR
XX (TOOH) UNITV QUEENS KINGSTON.
XX
XX Davies PL, Graham LA, Liou Y, Walker VK;
XX
XX WPI: 1999-095739/08.
DR
XX
XX New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage
XX
XX Claim 14; Page 56; 88pp; English.
XX
XX The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kb;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)
CC (1) binds specifically to antibodies raised against the Tenebrio molitor
CC YLI-YLA antifreeze proteins; (ii) is at least 60% identical with YLI-4
CC or (iii) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXAXXKXAXT (I). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific

```

Oy      21  CTC-----TCYCHCTGAD-----CTSCCTDAC-----TGGCN 48
          |||      |||      |||      |||      |||      |||
Db      48  CTCGASDRCSC-CAGGDBPHCSAAGDSCCTCAGSCKCKEECTCTSCCKSSCCSCCPVGCAC 106

```

```

QY 49 CPNAHTCTDSK-----NCVKA-----TCTGSTRKNTARTCTNSK----- 83
D 107 CAQCTCKGASDRKSCCAGGADPNCSCAAGDSCTCKASCKCKEKC-CTSCKKCCSCCP 165
QY 84 -----DCEFAKICTDSTNC---YKATPACTNS--TG 108
D 166 GCACCAQAGCTCKGASDRKSCCAGGADPNCSCAAGDSCTCKASCKCKEKCCTCKSCCSC 225
QY 109 CP 110
D 226 CP 227

RESULT 13
AAR05222
ID AAR05222 standard; protein; 2189 AA.
AC AAR05222:
DT 02-AUG-1990 (first entry)
DE Antigen GX5401FL encoded by Eimeria tenella genomic DNA.
KM Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
XX
OS Eimeria tenella.
PN MO9000403-A.
PD 25-JAN-1990.
PE 05-JUL-1989; 89WO-US02918.
PR 05-JUL-1988; 88US-0215162.
XX
PA (GENE-) GENEX CORP.
PI Anderson DM, McCandliss RJ, Strausberg SL, Strausberg RL;
DR WPI; 1990-051586/07.
XX N-PSDB; AA003324.
PT Cloned gene or fragment encoding antigenic protein -
PT which binds with antibodies against avian coccidia, and
PT transformed cells used in vaccine
XX
PS Claim 10; Page 93; Fig 14; 134pp; English.
XX
CC It is encoded by an open reading frame contained within the sequence of
CC clone 533 which was derived from an E. tenella genomic library screened
CC with radioactively labelled cDNA encoding the GX5401 antigen. It is of
CC about 250 kd. It carries several repeated peptide sequences and
CC is rich in cysteine residues. The open reading frame also encodes a
CC potential signal sequence for protein secretion. Also new are an
CC expression vector contg. cloned gene, and host cells transformed with
CC the vector. The transformed cells are used in a vaccine to immunise
CC birds against avian coccidiosis. By labelling the peptides, they can be
CC used as a type-specific probe. May also be used in an assay to detect
CC Ab against the coccidia. The Abs are used to identify transformed cells
CC contg. the DNA.
XX
SQ Sequence 2189 AA:

Query Match 17.8%; Score 118.5; DB 11; Length 2189;
Best Local Similarity 25.4%; Pred. No. 0.21;
Matches 32; Conservative 15; Mismatches 42; Indels 37; Gaps 6;

QY 20 MCLCTBCTC-----HCTGGADC---TSCDTACTG--CGNCPNAHTC-----TDSKN 60
D 537 ICTCTAGSGSDGTAGCHDCDDIDECLAENDCTPADGIGICENTVGSYTCACAGYQDGN 596
QY 61 CVKAATC-TGSTRKNTARTCTNSKDFEA-----KCTCTSTNCKYKATA 102

```

```

D 597 CDDIDECANGTINCLASATCTTQGSFECACNAGRSGNGVECNVDYDECTPADDCGNTL 656
QY 103 CTNSTG 108
D 657 CNIITVG 662

RESULT 14
AAB18144
ID AAB18144 standard; Protein; 1700 AA.
AC AAB18144:
DT 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:1.
KM Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
PN WO200025728-A2.
PD 11-MAY-2000.
PE 05-NOV-1999; 99WO-US26796.
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOEF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 29-33; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic life cycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 1700 AA:

Query Match 17.5%; Score 116; DB 21; Length 1700;
Best Local Similarity 29.2%; Pred. No. 0.27;
Matches 40; Conservative 4; Mismatches 53; Indels 40; Gaps 8;

```


Job time : 45.6471 secs

```
OY      5  TCGFSKKMLVIAVIMC-----LCTEYCHCTGGADCTSCDTA-----CTGC 46
      11  :      |      |      |      |      |      |      |      |
Db      636 TCAAAAGCAACATATCCAAATGAATTAACATCTGAGAGATCAACAGAAATTAATCCATGC 695
OY      47  GNCPPAHCTDTSKNCVKAATCTGSKCNTARTC-----TNSKDCFBAKTCYDSTNCY--- 98
      11  :      |      |      |      |      |      |      |      |
Db      696 -----AAACT---TC--AATATGATTAATACTAATGTTACTCATGTTTGGTCAAGA 745
OY      99  -KATACTNSTG---CPG 111
      11  :      |      |      |      |      |      |      |
Db      746 GTATCCTTGTGAACGG 762
```

RESULT 15

AAB36084 ID AAB36084 standard; Protein: 225 AA.

AAB36084;

16-FEB-2001 (first entry)

Neospora caninum NCP20 antigen.

Neospora caninum; NCP20; antiparasitic; vaccine; antigen;

neoparasitosis.

Neospora caninum.

W0200063244-A1.

26-OCT-2000.

20-APR-2000; 2000WO-AU00354.

21-APR-1999; 99AU-000928.

(INSE-) INSEARCH LTD.

Ellis JT, Atkinson R, Ryce CL;

WPI; 2000-679588/66.

N-PSDB; AAC67872.

Polynucleotide encoding parasite antigens and the antigenic polypeptide
useful for immunizing mammals against neosporosis -

Claim 4; Page 35-36; 39pp; English.

The present sequence is Neospora caninum antigen NCP20. A
composition comprising NCP20 is useful for obtaining a protective effect
against neosporosis in a mammal. The NCP20 polynucleotide may be
administered in the form of a DNA vaccine for vaccinating animals against
neoparasitosis. Antibodies against the NCP20 antigen may be identified
using the NCP20 polypeptide. NCP20 and antibodies specific to NCP20 are
useful for diagnosis of neosporosis.

SQ Sequence 225 AA;

Query Match 16.7%; Score 111; DB 21; Length 225;

Best Local Similarity 32.5%; Pred. No. 0.11;

Matches 27; Conservative 3; Mismatches 43; Indels 10; Gaps 3;

```
OY      30  CTGAGDCTCTDCTACGCGNCPNAHTCTDSKNCVKAATCTGSKCN-TARTCTNSKCFEA 88
      11  :      |      |      |      |      |      |      |      |
Db      18  CTTGTGCGCAAAACGACGCC-----CCTCCGGCTACCCGTTTGGCTAGTCT---CTAT 68
OY      89  KTCTDSTNCYKATACTNSTGCPG 111
      11  :      |      |      |      |      |      |      |
Db      69  TCCAGTAGGCCGTTCTGAATCCG 91
```

Search completed: October 21, 2002, 16:38:24

RESULT 7
US-08-977-767-3

Sequence 3, Application US/08977767
Patent No. 5972684
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1532042
US-08-977-767-3

Query Match 24.2%; Score 160.5; DB 2; Length 1345;
Best Local Similarity 38.3%; Pred. No. 9e-07;
Matches 41; Conservative 2; Mismatches 45; Indels 19; Gaps 6;

QY 16 AVTVMCLCTEYCHCTGADCTSC--TDACTGCGNCPNAHTCTDSKNCVKAA---TCTGS 70
DB 1012 AATATACACCTCCCTCCAG--CTGCATTCCCTGAGACT---TCTGAGCCGAGATCTCTCGG 1066

QY 71 TKCNTARTCTNSNDCEATCTDSTNCTYAT-----ACTNSTGCPG 111
DB 1067 GTCCAGCTCT---CCAGAGCTCAGCCGTAAAGCGCCGCTCGAG 1110

RESULT 8
US-08-508-761B-22
Sequence 22, Application US/08508761B
GENERAL INFORMATION:
APPLICANT: Joliff, Gwennael
APPLICANT: Guyonvarch, Arnel
APPLICANT: Purification, Relano
APPLICANT: Duchilton, Francis
APPLICANT: Renaud, Michel
TITLE OF INVENTION: System for Protein Expression and
TITLE OF INVENTION: Secretion Especially in Corynebacteria
NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh St. N.W.
CITY: Washington D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,761B
FILING DATE: 31-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09652
FILING DATE: 29-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09870
FILING DATE: 02-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P58525NA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: synthetic
US-08-508-761B-22

Query Match 24.0%; Score 159.5; DB 3; Length 120;
Best Local Similarity 39.1%; Pred. No. 1.3e-07;
Matches 34; Conservative 1; Mismatches 49; Indels 3; Gaps 3;

QY 23 CTCTCHCTGAGDCTCTDCTGCGNCPNAHTCTDSKNCVKAATCTGCTCNTARTCTNS 82
DB 36 CTGGCC-CTGGGCGCTGC-GCCTGTGCTCGAGCCCTGGGCGCTGGCGCTGCTGT 93

QY 83 KDFEAKTCTDSTNCTYATCTACTNSTGC 109
DB 94 GCTGGGCGCT-GAGCCTGTGCTGTGCTG 119

RESULT 9
US-08-900-230-3
Sequence 3, Application US/08900230
Patent No. 6329197
GENERAL INFORMATION:
APPLICANT: Baird, Jonathan A.
TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of The Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-C/JPM/ADM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-900-230-3

Query Match 23.7%; Score 157.5; DB 4; Length 1417;
Best Local Similarity 37.1%; Pred. No. 1.7e-06;
Matches 36; Conservative 0; Mismatches 50; Indels 11; Gaps 3;

QY 21 CLCTECYCHCT-----GGADCTSCDCTACTGCGNCPNAHTCTDSKNCVKATCTGSTRKN 74
DB 904 CTCACGGCGCTCTGCTGGGTCGCC--ACGACGGCGTCATCTCTGCG---TTCTGGTAGC 958
QY 75 TARTCTNSKDCFEAKTCTDSTNCKRATCTNSTGCPG 111
DB 959 GCCGCTTGCCTTCAGCCCGGCCACCTACGCTGTCGCC 995

RESULT 10
US-08-630-915A-37
Sequence 37, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOPMAN, No. 6309820H
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-37

Query Match 20.9%; Score 139; DB 4; Length 1400;
Best Local Similarity 29.7%; Pred. No. 7.3e-05;
Matches 38; Conservative 3; Mismatches 53; Indels 34; Gaps 5;

QY 16 AVIWMCLC-----TECYCHC-----TGADCTSCDCTACTGCGNCPNAHTCTDSK 59
DB 347 AAATCCCAAGAAATGAGGTCCCGCTCCAGTGAACCACTGATTCACATCTGCC 406
QY 60 NCVK-----AATCTG---STKNTAR-----TCTNSKDCFEAKTCTDSTNCKYAT 101
DB 407 CCTGCCCCCAACTGGCTTGGTAGACCCCCCGCCCTTGGCAGTAACTCTTCAGAG 466
QY 102 ACTNSTGC 109
DB 467 CCCTCCAC 474

RESULT 11
US-07-906-349A-6
Sequence 6, Application US/07906349A
Patent No. 5434064
GENERAL INFORMATION:
APPLICANT: Schlessinger, Joseph
APPLICANT: Skolnik, Edward Y.
APPLICANT: Margolis, Benjamin L.
TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES
TITLE OF INVENTION: TARGET PROTEINS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906,349A
FILING DATE: 30-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/643,237
FILING DATE: 18-JAN-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-349A-6
Query Match 19.7%; Score 130.5; DB 1; Length 801;
Best Local Similarity 35.1%; Pred. No. 0.00025;
Matches 34; Conservative 3; Mismatches 39; Indels 21; Gaps 6;
QY 23 CTCTCYCHCTG-----GADCTSCDCTACTGCGNCPNAHTCTDSKNCVKATCTGSTRKNAR 77

Db 556 CTTTCAGAAATATGTTTAATCATCATGCTGCTC-----TC-----CTCTTCCCTCCCTTTG 606
QY 78 TCTNSKDCFEAKTC--IDSTNCYKATACTNSTG-CPG 111
Db 607 TCTTTTTCATCCTTTTTC-----TCTTCTGTCAG 639

RESULT 12

US-08-990-823-112
; Sequence 112, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; CURRENT FILING DATE: 1997-12-15
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; EARLIER FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 112
; LENGTH: 143
; TYPE: PRN
; ORGANISM: Mycobacterium tuberculosis
US-08-990-823-112

Query Match

Best Local Similarity 17.6%; Score 117; DB 4; Length 143;
Matches 36; Conservative 2; Mismatches 47; Indels 40; Gaps 8;

QY 21 CLCTEC-----YC--HCTGAGDC-----TSCCTACTGCGC-----NCPNAHTCTD 57
Db 17 CACACAGAGTGTACGACGAGACCCCGCCAGCTTGGCGACGCGGAATCTTCCGCTG 76
QY 58 SKNCVKATCTGSTRKNT-----ARTCTNSKDCFEAKTCTDSTNCYKATACTN-----S 106
Db 77 TGGC-----TCACGACCTTTCGCCGCACTTCAC--GATCTTGACATGACGCGCAGAA 130
QY 107 TGCGP 111
Db 131 CGCGC 135

RESULT 13

US-08-974-022-53
; Sequence 53, Application US/08974022
; Patent No. 6015938
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,022

; FILING DATE: 12-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-974-022-53

Query Match

Best Local Similarity 17.4%; Score 115.5; DB 3; Length 102;
Matches 35; Conservative 2; Mismatches 38; Indels 25; Gaps 7;

QY 23 CTECYCHCTGGADCTISCDTACTG-----CGMCPNAHTCTDSKNCVKATC--TGST 71
Db 15 CTTTCATCAG--CTGCTGTGTGATTAATGTCGCCGGGTACA--CCTACTTCTTTGAA 69
QY 72 KCTARTCTNSKDCFEAKTCTDSTNCYKATACTNSTGCGP 111
Db 70 GAGTAGTCGACGAC--ACACTATT-----TACAG--GC GG 100

RESULT 14

US-08-795-445A-53
; Sequence 53, Application US/08795445A
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-445A-53

Query Match

Best Local Similarity 17.4%; Score 115.5; DB 4; Length 102;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:36:38 ; Search time 22.2353 seconds
(without alignments)
484.006 Million cell updates/sec

Title: US-10-032-658-11

Perfect score: 664
Sequence: 1 MAFKTCGFSSKMWLVIAIVM.....DSTNCKYKATCTNSTGCPGH 112

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138.5	20.9	313	2	extracellular matr
2	128.5	19.4	1046	2	prestalk protein p
3	119.5	18.0	861	2	Notch homolog Motc
4	118.8	17.8	4660	2	9p30 protein prec
5	117.5	17.7	2531	2	Notch protein homo
6	117.5	17.7	2531	2	Notch-1 protein -
7	114	17.2	2555	2	Notch protein homo
8	113.5	17.1	548	2	hypothetical prote
9	113.5	17.1	3002	2	fibritillin 1 precu
10	113	17.0	2524	2	Xotch protein - Af
11	112.5	16.9	169	1	ultra high-sulfur
12	112	16.9	1111	2	hypothetical prote
13	111	16.7	677	2	trophocyste cystei
14	111	16.7	1101	2	hypothetical prote
15	110.5	16.6	2871	2	fibritillin I - bovi
16	109	16.4	230	2	ultra-high-sulfur
17	109	16.4	798	2	Integrin beta-7 ch
18	108.5	16.3	2871	2	fibritillin-1 precu
19	108	16.3	640	1	uromodulin precurs
20	107.5	16.2	188	2	high sulfur protei
21	107.5	16.2	1700	2	Balbani ring 3 pr
22	107.5	16.2	2703	1	notch protein - Fr
23	107	16.1	601	2	DIF-induced presta
24	107	16.1	1712	2	masking protein pr
25	106.5	16.1	2352	2	Notch homolog prot
26	106.5	16.0	1220	2	jagged protein prote
27	106.5	16.0	3507	2	hypothetical prote
28	106	16.0	223	2	ultra-high-sulfur
29	106	16.0	738	2	hypothetical prote

30	106	16.0	739	2	B88553	protein K04H4.2b [
31	105.5	15.9	574	2	B88465	protein B0244.8 [1
32	105	15.7	674	2	I55476	growth potentiati
33	104	15.7	2825	2	T14271	Doc4 protein, stre
34	103.5	15.6	58	2	S43367	metallothionein
35	103.5	15.6	65	2	A38739	metallothionein -
36	103.5	15.6	680	2	PN0510	Integrin beta-3 ch
37	103.5	15.6	1299	2	T43251	furin (EC 3.4.21.7
38	103.5	15.6	1895	2	T15881	hypothetical prote
39	103.5	15.6	2531	2	T31070	notch homolog - se
40	103	15.5	753	2	B36268	platelet glycoprot
41	103	15.5	778	2	A60798	platelet glycoprot
42	103	15.5	788	2	A26547	platelet glycoprot
43	103	15.5	788	2	T17349	platelet glycoprot
44	103	15.5	1557	2	T28811	hypothetical prote
45	103	15.5	2946	2	T15840	hypothetical prote

ALIGNMENTS

RESULT 1
S44208
extracellular matrix protein B - Dictyostelium minutum (fragment)
C:Species: Dictyostelium minutum
C:Date: 07-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 05-Nov-1999
C:Accession: S44208
R:van Es, S.; Nieuwenhuisen, B.W.; Lenouvel, F.; van Deursen, E.M.; Schaap, P.
Submitted to the EMBL Data Library, April 1994
A:Description: Universal signals control slime mold stalk information.
A:Reference number: S44208
A:Accession: S44208
A:Molecule type: DNA
A:Residues: 1-313 <V>
A:Cross-references: EMBL:X76948; NID:9474918; PIDN:CAA55545.1; PID:9474919
C:Genetics:
A:Introns: 24/1
C:Keywords: extracellular matrix

Query Match 20.9%; Score 138.5; DB 2; Length 313;
Best Local Similarity 29.6%; Pred. No. 0.0015;
Matches 32; Conservative 9; Mismatches 46; Indels 21; Gaps 3;

QY 23 CTCTCTCTGGA-----DCTCT-DACGTGCGCPNAHCTPSKNCVKATCTGSKTC 73
DB 91 CTLDSCDCKGCGCTHPMNCBDHNRCTLDSCDCKGCGCTHPMNCDDKNACTVDSGNSSTGC 150

QY 74 N-----TARTCTNSKDCFEAKTCPTDSTNCKYKATCTNSTGC 109
DB 151 SHTPTSCDYNSTVDSCDNCGCVHTPTACDDKNACTVDSGNSKTCG 198

RESULT 2
A26838
prestalk protein precursor - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 22-Nov-1996
C:Accession: A26838
R:Ceccarelli, A.; McRebbbie, S.J.; Jernyn, K.A.; Duffy, K.; Early, A.; Williams, J.G.
Nucleic Acids Res. 15; 7463-7476, 1987
A:Title: Structural and functional characterization of a Dictyostelium gene encoding
A:Reference number: A26838; MUID:88015608
A:Accession: A26838
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1046 <CEC>
A:Experimental source: Strain Ax-2
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1046/Product: prestalk protein #status predicted <MAT>

Query Match 19.4%; Score 128.5; DB 2; Length 1046;
Best Local Similarity 35.9%; Pred. No. 0.015;
Matches 33; Conservative 9; Mismatches 37; Indels 13; Gaps 5;

QY	23	CTCEYCHCTGAGDGT--SCRD--ACGCGCNCNAHCTCTSKNCVYA-ATCGTGTCTTAR	77
Db	800	CTEKCCTQNGSVYHTPIRCDDLNSCI-----ADCSNSTGCVHTPIICDDNNKC-TAD	851
QY	78	TCTNKAOCFEAKTCTDSTNCYKATFACTNSTGC	109
Db	852	SCSNSTGCGCHTPIISCDNNPNCTVDSCSNSTGC	883

RESULT 3

Notch homolog Match protein - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
 C:Accession: A48825
 R:Resume, A.G.; Conlon, R.A.; Ziringibl, R.; Yamaguchi, T.P.; Rossant, J.
 Dev. Biol. 154, 377-387, 1992
 A:Title: Expression analysis of a Notch homologue in the mouse embryo.
 A:Reference number: A48825; MUID:93050801
 A:Accession: A48825
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-861 <RG>
 A:Experimental source: embryo
 A:Note: sequence extracted from NCBI backbone (NCBIP:119144)
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology;
 F:26-57/Domain: EGF homology <EGF>
 F:64-95/Domain: EGF homology <EGX1>
 F:196-229/Domain: EGF homology <EGF2>
 F:441-472/Domain: EGF homology <EGX2>

Query Match	18.0%	Score 119.5;	DB 2;	Length 861;
Best Local Similarity	32.7%	Pred. No. 0.062;		
Matches 36;	Conservative	7;	Mismatches 50;	Indels 17;
			Gaps	6;
QY	15	IAYVMC----	LCTE----	CYCHCTGAGDICTACTGC--GNCPSNAHCTD--SKNC 61
		: :	: :	: :
Db	318	IDVTLLQHGGLVYDSDGKHCHCQAGTGTGTCEDVEDCSFNCQAGATDIDLGFSG 377		
		: :	: :	: :
QY	62	VKATCTGCTKCTWARTCTNSKDFEAKTCTDSTCTCATCTNSTGCPG 111		
		: :	: :	: :
Db	378	KCVAGHYGNSCSEINECL-SQPCQNGGTCTDITNSTYK---CSAPRGTGG 423		

RESULT 4

gp330 protein precursor - rat
N:Alternate names: megalin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-2000 #sequence-revision 11-Jan-2000 #text-change 04-Mar-2000
C:Accession: J42727
R:Salto, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A:Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of the gp330 family
A:Reference number: A58173; MUID:95024033
A:Accession: J42737
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-4660 <SNT>
A:Cross-references: EMBL:L34049; NID:9561852; PID:9561853; PLDN:AA51369.1
A:Experimental source: strain Sprague-Dawley; kidney
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
F:1-25/Domain: signal sequence #status predicted <SNT>
F:26-4660/Product: gp330 protein #status predicted <MAT>

```

Query Match 17.8% Score 118; DB 2; Length 4660;
Best Local Similarity 23.7%; Pred. No. 0.21;
Matches 36; Conservative 14; Mismatches 46; Indels 56; Gaps 8;

12 WLVIANYVWCL-----CTEYC-----HCTGADCTCTD-----41
10 WMLLAIACLEVVSSQEGSGNFRCDNGYCIIPASWRDGDTRDLDTDITGCPSPSCS 69
Ob

```

```

OY  42  ---ACTGGGNC-PAHAACTDTSKNCVKA-----TCGSGCKTATCTNSK-----DCFE  87
Db   70  GLFLCPACGCTCLPSSWVCDEKDCDGDAGDEQGNACGTTCSNQMTCSNGCCIPSEYICDH  129
OY   88  AKTC---TDSNCKYAT-----ACTNST  107
Db  130  VSDCPDGSDEBNCHPCTDQLTCANGAACYYNYS  161

```

RESULT 5

Notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
 C:Accession: S18188
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S18188; MUID:92111383
 A:Accession: S18188
 A:Molecule type: mRNA
 A:Residues: 1-2531 <MEI>
 A:Cross-references: EMBL:X57405; NID:957634; PID:957635
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:1987-1018/Domain: EGF homology <EGF1>
 F:1025-1056/Domain: EGF homology <EGF>
 F:1233-1264/Domain: EGF homology <EGF2>
 F:1917-1949/Domain: ankyrin repeat homology <AN1>
 F:1950-1982/Domain: ankyrin repeat homology <AN2>
 F:1984-2016/Domain: ankyrin repeat homology <AN3>
 F:2017-2049/Domain: ankyrin repeat homology <AN4>
 F:2050-2082/Domain: ankyrin repeat homology <AN5>

[illegible]

RESULT 6

Notch-1 protein - mouse
N:Alternate names: notch protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A46019; S25144
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
Genomics 15, 259-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
A:Reference number: A46019; MUID:93194170
A:Accession: A46019
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531
A:Cross-references: GB:S11886; GB:S47228; NID:q288502; PIDN:CAA77941.1; PID:q288503
A:Note: sequence extracted from NCBI backbone (NCBI:P127318)
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.
submitted to the EMBL Data Library, April 1992
A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, sugg
A:Reference number: S25144
A:Accession: S25144
A:Molecule type: mRNA
A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
A:Cross-references: EMBL:Z11886
A:Genetics:

A:Gene: notch-1
 A:Map position: 2
 A>Note: Proximal region of chromosome 2
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:106-133/Domain: EGF homology <EGF1>
 F:144-175/Domain: EGF homology <EGF1>
 F:222-254/Domain: EGF homology <EGF2>
 F:261-292/Domain: EGF homology <EGF2>
 F:339-370/Domain: EGF homology <EGF3>
 F:416-449/Domain: EGF homology <EGF3>
 F:456-487/Domain: EGF homology <EGF4>
 F:494-525/Domain: EGF homology <EGF5>
 F:532-563/Domain: EGF homology <EGF6>
 F:607-638/Domain: EGF homology <EGF7>
 F:682-713/Domain: EGF homology <EGF8>
 F:757-788/Domain: EGF homology <EGF9>
 F:795-826/Domain: EGF homology <EGF10>
 F:873-904/Domain: EGF homology <EGF11>
 F:911-942/Domain: EGF homology <EGF12>
 F:949-980/Domain: EGF homology <EGF13>
 F:987-1018/Domain: EGF homology <EGF14>
 F:1025-1056/Domain: EGF homology <EGF15>
 F:1063-1094/Domain: EGF homology <EGF16>
 F:1149-1180/Domain: EGF homology <EGF17>
 F:1187-1218/Domain: EGF homology <EGF18>
 F:1233-1264/Domain: EGF homology <EGF19>
 F:1352-1383/Domain: EGF homology <EGF19>
 F:1391-1425/Domain: EGF homology <EGF>
 F:1917-1948/Domain: ankyrin repeat homology <AN1>
 F:1949-1981/Domain: ankyrin repeat homology <AN2>
 F:1983-2015/Domain: ankyrin repeat homology <AN3>
 F:2016-2048/Domain: ankyrin repeat homology <AN4>
 F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 17.7% Score 117.5; DB 2; Length 2531;
 Best Local Similarity 32.1%; Pred. No. 0.16;
 Matches 35; Conservative 7; Mismatches 46; Indels 21; Gaps 6;
 QY 15 IAVIWC-----LCTE-----CYCHCTGAGACTCTGCTACTGC--GNCNPAHTCTD---SKNC 61
 DB 1110 IDTLLCQHGGLCVDEGDNRHYCHCQAGYGYSCYCEDEVDSCSPNQNGATCTDYLGFSFC 1169
 QY 62 VKAATCGSTKCTNARTCTNSKDCFEAKTCTDSTNCYKATCTNSTGCP 110
 DB 1170 KCVAGHGSNCSEINECL-SQPCQNGTCTIDLTNSYKCS-----CP 1210

RESULT 7
 A40043
 notch protein homolog TAN-1 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 13-Aug-1999
 C:Accession: A40043
 R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, Cell 66, 649-661, 1991
 A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal
 A:Reference number: A40043; MUID:91347367
 A:Accession: A40043
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2555 <ELL>
 A:Cross-references: GB:M73980
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:261-292/Domain: EGF homology <EGF1>
 F:494-525/Domain: EGF homology <EGF1>
 F:987-1018/Domain: EGF homology <EGF2>
 F:1149-1180/Domain: EGF homology <EGF2>
 F:1187-1218/Domain: EGF homology <EGF3>
 F:1233-1264/Domain: EGF homology <EGF3>
 F:1927-1959/Domain: ankyrin repeat homology <AN1>
 F:1960-1992/Domain: ankyrin repeat homology <AN2>
 F:1994-2026/Domain: ankyrin repeat homology <AN3>
 F:2027-2059/Domain: ankyrin repeat homology <AN4>

F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 17.2% Score 114; DB 2; Length 2555;
 Best Local Similarity 31.6%; Pred. No. 0.29;
 Matches 31; Conservative 6; Mismatches 45; Indels 16; Gaps 6;
 QY 23 CTE-----CYCHCTGAGACTCTGCTACTGCN--CPNPAHTCTDSKNCYKAATCT-----GST 71
 DB 922 CTDSINAFDCDLPFGFGTCEEDINECASDPKRNAGNCID---CYDSTYCTCPAFPSGI 978
 QY 72 KC-NTARTCTNSKDCFEAKTCTDSTNCYKATCTNSTG 108
 DB 979 HCENNTPDCIES-SCFNGTCTVDGINSTCTCLCPGFG 1015

RESULT 8
 T16642
 hypothetical protein R01H2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16642
 R:Babbs, N.
 submitted to the EMBL Data Library, February 1994
 A:Description: The sequence of C. elegans cosmid R01H2.
 A:Reference number: Z18553
 A:Accession: T16642
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-548 <BAB>
 A:Cross-references: EMBL:U00035; NID:9458989; PID:9458992; PIDN:AAA50645.1; CESP:R01H
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:R01H2.3
 A:Introns: 18/2; 44/3; 259/3; 495/1

Query Match 17.1% Score 113.5; DB 2; Length 548;
 Best Local Similarity 25.0%; Pred. No. 0.13;
 Matches 37; Conservative 14; Mismatches 36; Indels 61; Gaps 9;
 QY 19 VMCLCTECYC-----HCTGGADCTSC-----TDA-----CTGCNCNCP--- 51
 DB 188 LITCLAEKLCGVNCPDGSDEASCRSKSGKDPKSGSDACLPISVKGCVSPSCNENSD 247
 QY 52 -----AHTCTDSKNCVKAA-TCTGSTKCN-----TARTCTNSKDCFEAKTCT 92
 DB 248 ESNCKCKGKAHNC--GRNCTKASKVCDGIPDCDDGSDERHQCCKTSGS-----ERALCE 301
 QY 93 DSTNCKATAC-----TNSTGCPG 111
 DB 302 DGTCLMRSGQVCDGKHNDCLDIDEENCPG 329

RESULT 9
 A47221
 fibrillin 1 precursor - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
 C:Accession: A47221; I54355; S17064; I59574; S17062; S62111; A34198
 R:Corson, G.M.; Chalberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y. Genomics 17, 476-484, 1993
 A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain stru
 A:Reference number: A47221; MUID:94010947
 A:Accession: A47221
 A:Molecule type: mRNA
 A:Residues: 1-337; T, 339-1029 <COR>
 A:Cross-references: GB:X63556
 R:Perleira, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Pangillan, T.; Bo Hum. Mol. Genet. 2, 961-968, 1993
 A:Title: Genomic organization of the sequence coding for fibrillin, the defective gen
 A:Reference number: I54355; MUID:93372860
 A:Accession: I54355
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

A:Residues: 132-3002 <PER>
A:Cross-references: GB:LI3923; NID:g306745; PIDN:AA02036.1; PID:g306746
R:Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.
Nature 352, 334-337, 1991
A:Title: Partial sequence of a candidate gene for the Marfan syndrome.
A:Reference number: S17064; MUID:91304568
A:Accession: S17064
A:Molecule type: mRNA
A:Residues: 1030-3002 <MAS>
A:Cross-references: EMBL:X63556
R:Diez, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.
Science 259, 680-683, 1993
A:Title: The skipping of constitutive exons in vivo induced by nonsense mutations.
A:Reference number: I59574; MUID:93157831
A:Accession: I59574
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 2217-2288, '1', 2290-2325 <RES>
A:Cross-references: GB:S54426; NID:g264860; PIDN:AA025244.1; PID:g264861
R:Lee, B.; Godfrey, M.; Vitale, E.; Horl, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras, P.
Nature 352, 330-334, 1991
A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different
A:Reference number: S17062; MUID:91304567
A:Accession: S17062
A:Molecule type: mRNA
A:Residues: 'VLVTVFPLSYKNML', 944-1444 <LEE1>
A:Cross-references: EMBL:X62008; NID:g31398; PIDN:CA056534.1; PID:g5924015
A:Accession: S62111
A:Molecule type: protein
A:Residues: 1166-1176, 'X', 1178-1180, 'D', 1182-1185 <LEE2>
R:Madol, K.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.
J. Biol. Chem. 264, 21381-21385, 1989
A:Title: Connective tissue microfibrils. Isolation and characterization of three large
A:Reference number: A34198; MUID:90078246
A:Accession: A34198
A:Molecule type: protein
A:Residues: 565-575; 1890-1892, 'T', 1894-1900 <MAD>
C:Comment: Fibrillin is a major component of elastin-associated microfibrils.
C:Genetics:
A:Gene: GDB:FBNI
A:Cross-references: GDB:127115; OMIM:134797; OMIM:154700
A:Map position: 15q21.1-15q21.1
A:Introns: 2236/1; 2258/1; 2297/1
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; N
F:1-3002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted
F:132-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MATC>
F:132-1367/Domain: EGF homology <EGF>
F:1457-1492/Domain: EGF homology <EGF2>
F:2262-2295/Domain: EGF homology <EGF1>
Query Match 17.1%; Score 113.5; DB 2; Length 3002;
Best Local Similarity 27.1%; Pred. No. 0.34; Mismatches 41; Indels 47; Gaps 7;
Matches 36; Conservative 9;
OY 23 CT-----ECYCH-----CTGAGDCTCTDCTG--CGCNPNAHTC----- 55
DB 1343 CTNSEGSEYECSCQGFALMPDQRCSTDIDECEDNPNICDGGGCTNINIGEVRLCYDGFMA 1402
OY 56 -TDSKNCKKATCTGSTRKCNARTCTNSKOCF-----EAKT-CTD-----ST 95
DB 1403 SEDKKTCDVNECDLNPICLSTGTCENTKGSFICHCMDMGYSGKKGKTCYDINECEIGAH 1462
OY 96 NCYKATACNTSTG 108
DB 1463 NCGKHAVCNTNAG 1475
RESULT 10
A35844
Xotch protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 13-Aug-1999

C:Accession: A35844
R:Coffman, C.; Harris, W.; Kintner, C.
Science 249, 1438-1441, 1990
A:Title: Xotch, the Xenopus homolog of Drosophila notch.
A:Reference number: A35844; MUID:90385285
A:Accession: A35844
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: mRNA
A:Residues: 1-2524 <COR>
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
C:Keywords: transmembrane protein
F:146-177/Domain: EGF homology <EGX1>
F:184-215/Domain: EGF homology <EGF1>
F:222-254/Domain: EGF homology <EGF>
F:456-487/Domain: EGF homology <EGX2>
F:757-788/Domain: EGF homology <EGF3>
F:1025-1056/Domain: EGF homology <EGX3>
F:1924-1956/Domain: ankyrin repeat homology <AN1>
F:1957-1989/Domain: ankyrin repeat homology <AN2>
F:1991-2023/Domain: ankyrin repeat homology <AN3>
F:2024-2056/Domain: ankyrin repeat homology <AN4>
F:2057-2089/Domain: ankyrin repeat homology <AN5>
Query Match 17.0%; Score 113; DB 2; Length 2524;
Best Local Similarity 34.3%; Pred. No. 0.34; Mismatches 42; Indels 18; Gaps 6;
Matches 34; Conservative 5;
OY 23 CTCTCY----CHCT---GGADCTCTDCTGCGCNPNAHTCTDSKNCKKATC-----TG 69
DB 960 CTCCTCAVSYTCCTCPGPGSGHICSESNPDCYE--SSCFNGGTCIDG---INTFTCCPPGFTG 1015
OY 70 STCKNTARCTNSKDCFEAKTCGDSNCKKATACNTSTG 108
DB 1016 STCGHDINEC-DSKPCPLNGTCDSTGYTKCTCPGCTG 1053
RESULT 11
S18946
ultra high-sulfur keratin 1 - human
N:Alternate names: UHS keratin; ultra high-sulfur matrix protein
C:Species: Homo sapiens (man)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: S18946; B36686
R:Drabent, B.; Doecke, D.
submitted to the EMBL data library, December 1991
A:Description: Nucleotide sequence of a human high-sulphur keratin cDNA.
A:Reference number: S18946
A:Accession: S18946
A:Molecule type: mRNA
A:Residues: 1-169 <DRA>
A:Cross-references: EMBL:X63755; NID:g32471; PIDN:CAA45283.1; PID:g32472
R:MacKinnon, P.T.; Powell, B.C.; Rogers, G.E.
J. Cell Biol. 111, 2587-2600, 1990
A:Title: Structure and expression of genes for a class of cysteine-rich proteins of t
A:Reference number: A3686; MUID:91115951
A:Accession: A3686
A:Molecule type: DNA
A:Residues: 1-39, 'Y', 41-169 <MAC>
A:Cross-references: GB:X55293; NID:g34078; PIDN:CAA39005.1; PID:g34079
C:Genetics:
A:Gene: GDB:KRNI
A:Cross-references: GDB:125257; OMIM:148021
A:Map position: 11q13-11q13
C:Superfamily: ultra-high-sulfur keratin
C:Keywords: hair; tandem repeat
F:7-15/Region: Ser-rich nonapeptide repeat
F:59-68/Region: Gly-rich decapeptide repeat
F:69-78/Region: Gly-rich decapeptide repeat
F:79-88/Region: Cys-rich decapeptide repeat
F:89-97/Region: Ser-rich nonapeptide repeat
F:98-107/Region: Cys-rich decapeptide repeat
F:108-117/Region: Cys-rich decapeptide repeat
F:118-126/Region: Ser-rich nonapeptide repeat

F:127-136/Region: Cys-rich decapeptide repeat
 F:137-145/Region: Ser-rich nonapeptide repeat
 F:146-155/Region: Cys-rich decapeptide repeat
 F:156-165/Region: Cys-rich decapeptide repeat

Query Match 16.9%; Score 112.5; DB 1; Length 169;
 Best Local Similarity 26.5%; Pred. No. 0.081;
 Matches 27; Conservative 14; Mismatches 44; Indels 17; Gaps 4;
 Oy 19 VMCLCTEYCHCTGADCTCTDCTAGCGNCPNNAHTCTDSNCKATCTGSG-----70
 Db 45 VCCCVAPACSSSCGKRGCGSGSGSGSC-GCGSCSCCKPCCSSGCGSCCGSCCK 103
 Oy 71 ---TKNTARTCTNSKDFEAKTCTDSTNCKATCTNSTGC 109
 Db 104 PYCSGCSCKKPCSSSG--RSSSCGSSC---KPCSSSGC 140

RESULT 12
 T26972
 hypothetical protein Y47H9C.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T26972
 R:Harris, B.
 submitted to the EMBL Data Library, October 1998
 A:Reference number: Z20293
 A:Accession: T26972
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1111 <WTL>
 A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4
 A:Experimental source: clone Y47H9C
 C:Genetics:
 A:Gene: CESP:Y47H9C.4
 A:Map position: 1
 A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 16.9%; Score 112; DB 2; Length 1111;
 Best Local Similarity 30.4%; Pred. No. 0.25;
 Matches 31; Conservative 4; Mismatches 43; Indels 24; Gaps 6;
 Oy 28 CHCTGADCTG-----CTDAGCGNCPNNAHTCTDSK---NCVKAATCTGCTKNTART 78
 Db 164 CDCENGANDPELCTCTSGFG-ERCEK--PCPDNKGNCVKSCCGCGKCKNEKG 220
 Oy 79 CTNS-----KDCFEAK--TCTDSTNCKATATCTNSTG 108
 Db 221 CVCSIDGSGEFCINCKCEGKFGAECKFCNCQNGATCQNTNG 262

RESULT 13
 C42125
 trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)
 N:Alternate names: CRP72
 C:Species: Giardia lamblia
 C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
 C:Accession: C42125
 R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
 Mol. Cell. Biol. 12, 1194-1201, 1992
 A>Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 gene
 A:Reference number: A42125; MUID:92186850
 A:Accession: C42125
 A:Molecule type: DNA
 A:Residues: 1-677 <ADA>
 A:Cross-references: GB:M83934; NID:g159133
 A:Experimental source: trophozoites
 A>Note: sequence extracted from NCBI backbone (NCBI:88443, NCBI:P:88444); this ORF is not
 C:Keywords: surface antigen

Query Match 16.7%; Score 111; DB 2; Length 677;
 Best Local Similarity 29.2%; Pred. No. 0.23;

Matches 35; Conservative 13; Mismatches 36; Indels 36; Gaps 8;
 Oy 22 LCTEYCHCTG-----GADCTGCT-----DAGCGNCPNNAHTCTDSKNCVAA 65
 Db 104 LCTECKT-ANGLEKNPAPEKSGECILSDINGAGDYGVAACAO--CTKSDSNKGAA 159
 Oy 66 TCT-----GSTKCN-TARTC-TNSKDCFEAKTCTDSTNCKATCTNSTGCPGH 112
 Db 160 TCTACGAGYKPKDPAQSKCDGTCTCTETSAQC---TSCPEGKYLKGDASCYNNNGCTGN 216

RESULT 14
 T16840
 hypothetical protein T10E10.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16840
 R:Geisels, C.
 submitted to the EMBL Data Library, October 1995
 A:Description: The sequence of C. elegans cosmid T10E10.
 A:Reference number: Z18588
 A:Accession: T16840
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1101 <GEI>
 A:Cross-references: EMBL:U39644; NID:g1049339; PID:g1049343; PIDN:AAA80360.1; CESP:T1
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:T10E10.4
 A:Introns: 93/2; 152/2; 191/3; 209/2; 283/3; 303/1; 399/3; 421/1; 440/1; 465/1; 547/3
 Query Match 16.7%; Score 111; DB 2; Length 1101;
 Best Local Similarity 25.4%; Pred. No. 0.3;
 Matches 30; Conservative 4; Mismatches 50; Indels 34; Gaps 5;

Oy 23 CTETCYCHCTGADCT-----TCTDAGCGNCPNNAHTCTDSKNC-----61
 Db 661 CGRIGVDCNNAGACPLPTCPNNAISSQRCSGGCTNC--CPVGQTCMNGCCDILPSPSGG 718
 Oy 62 VKAATCTGSKNTARTCTNSKDC-----FEAKTCTDSTNCKATATCTNSTGCP 110
 Db 719 FAISMCTG--KCGTGEPCVNSGCCSLPCPSGLMSVORCAMGIGCPPNGCENGVCPP 774

RESULT 15
 A55567
 fibrillin I - bovine
 C:Species: Bos primigenius taurus (cattle)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
 C:Accession: A55567
 R:Tilstra, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
 Genomics 23, 480-485, 1994
 A>Title: Sequence number of the coding region of the bovine fibrillin cDNA and localization
 A:Reference number: A55567; MUID:95137597
 A:Accession: A55567
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2871 <TII>
 A:Cross-references: GB:I28748; NID:g508427; PIDN:AAA74122.1; PID:g508428
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:1201-1236/Domain: EGF homology <EGF>

Query Match 16.6%; Score 110.5; DB 2; Length 2871;
 Best Local Similarity 26.3%; Pred. No. 0.55;
 Matches 35; Conservative 10; Mismatches 41; Indels 47; Gaps 7;

Oy 23 CT-----ECYCH-----CTGADCTGCTDAGTG--CGNCPNNAHTC-----55
 Db 1212 CTNSESGYECSCQPGFALPDPQRSCTDIDCEDNPNNICDGGQCTNIPGEYRCICYDGFMA 1271
 Oy 56 -TDSKNVKAATCTGSKNTARTCTNSKDCF-----EAKT-CTD-----ST 95
 Db 1272 SEDMKTCVDVNECDLNPNICLSGTCENTYKGFICHCMDMGYSGKRGKGTGCTDINECEIGAH 1331

OY 96 NCYKATACTNSTG 108
|| : || : |
Db 1332 NCDRRAVCTNTAG 1344

Search completed: October 21, 2002, 16:40:12
Job time : 25.2353 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:23 ; Search time 13.1765 Seconds
(without alignments)
329.116 Million cell updates/sec

Title: US-10-032-658-11
Perfect score: 664
Sequence: 1 MAFPTGCFSKMVLVIAYIVM.....DSTNCKYKATACNSTGCPGH 112

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128.5	19.4	1046	PSTA.DICDI	P11976 dictyostell
2	118	17.8	4660	LRP2_RAT	P81158 rattus norv
3	117.5	17.7	2531	NTC1_MOUSE	Q01705 mus musculu
4	117.5	17.7	2531	NTC1_RAT	Q07008 rattus norv
5	115	17.3	194	KRUB_HUMAN	Q75690 homo sapien
6	114	17.2	2444	NTC1_HUMAN	P46531 homo sapien
7	113.5	17.1	2871	FBN1_HUMAN	P35555 homo sapien
8	113	17.0	2524	NOTC2_XENLA	P21783 xenopus lae
9	112.5	16.9	169	KRUA_HUMAN	P26371 homo sapien
10	110.5	16.6	2871	FBN1_BOVIN	P81133 bos taurus
11	110.5	16.6	2871	FBN1_PIG	Q94736 sus scrofa
12	109	16.4	773	MT_DREPO	Q44550 dreissena p
13	109	16.4	798	ITB7_HUMAN	P26010 homo sapien
14	108.5	16.3	328	C170_GIALA	P41951 caenorhabdi
15	108.5	16.3	2871	FBN1_MOUSE	Q61554 mus musculu
16	108	16.3	640	URON_HUMAN	P07911 homo sapien
17	107.5	16.2	1700	BAR3_CHITE	Q03376 chironomus
18	107.5	16.2	2703	NOTC2_DROME	P07207 drosophila
19	107	16.1	1712	TGFB_RAT	Q00918 rattus norv
20	106	16.0	1246	YMW2_CAEEL	P43504 caenorhabdi
21	104.5	15.7	643	URON_MOUSE	P48733 bos taurus
22	104.5	15.7	787	ITB6_MOUSE	Q92015 mus musculu
23	104	15.7	1592	SORL_CHICK	Q98930 g sortilin-
24	103.5	15.6	58	MT_CARMA	P55948 garcinus ma
25	103.5	15.6	65	MTB_STRPU	Q27287 strongyloce
26	103.5	15.6	787	ITB3_MOUSE	Q54890 mus musculu
27	103.5	15.6	1895	YLK3_CAEEL	P41951 caenorhabdi
28	103	15.5	788	ITB3_HUMAN	P5106 homo sapien
29	103	15.5	1557	LMWL_CAEEL	Q18823 caenorhabdi
30	102.5	15.4	577	ITB6_CAYPO	P18563 cavia porce
31	102.5	15.4	909	LDL1_XENLA	Q99087 xenopus lae
32	102	15.4	644	URON_RAT	P27590 rattus norv
33	102	15.4	913	PKC5_HUMAN	Q92824 homo sapien

34	102	15.4	1178	1	TSP2_CHICK	P35440 gallus gall
35	101.5	15.3	755	1	COMP_RAT	P35444 rattus norv
36	101.5	15.3	860	1	LDLR_HUMAN	P01130 homo sapien
37	101.5	15.3	1064	1	FBP1_STRPU	P10079 strongyloce
38	101.5	15.3	1107	1	YLK2_CAEEL	P41950 caenorhabdi
39	101	15.2	667	1	TS11_GIALA	Q03185 giardia lam
40	101	15.2	1184	1	FBP2_HUMAN	P98095 homo sapien
41	100.5	15.1	798	1	ITB1_FELCA	P53713 felis silve
42	100	15.1	1221	1	FBP2_MOUSE	P37889 mus musculu
43	99	14.9	705	1	FBP1_MOUSE	Q08879 mus musculu
44	99	14.9	1173	1	TSPI_XENLA	P35448 xenopus lae
45	99	14.9	2437	1	NOTC_BRARE	P46530 brachydantio

ALIGNMENTS

RESULT 1	ID	PSTA.DICDI	STANDARD:	PRT: 1046 AA.
AC	P11976;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Prestalk protein precursor.			
GN	ECMB.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.			
OX	NCBI_Taxid=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AX2;			
RX	MEDLINE=88015608; PubMed=3658700;			
RA	Ceccarelli A., McRobbie S.J., Jermyn K.A., Duffy K., Early A., Williams J.G.;			
RT	"Structural and functional characterization of a Dictyostelium gene encoding a DIF inducible, prestalk-enriched mRNA sequence.";			
RL	Nucleic Acids Res. 15:7463-7476(1987).			
RN	[2]			
RP	SEQUENCE OF 1-17 FROM N.A.			
RX	MEDLINE=87257883; PubMed=3600646;			
RA	Ayres K., Neuman W., Rowekamp W.G., Chung S.;			
RT	"Developmental regulation of Phase I-hypersensitive sites in Dictyostelium discoideum.";			
RL	Mol. Cell. Biol. 7:1823-1829(1987).			
CC	-I- INDUCTION: BY THE PUTATIVE STALK-SPECIFIC MORPHOGEN DIF			
CC	(DIFFERENTIATION INDUCING FACTOR).			
CC	-I- SIMILARITY: THE REPEATED SEQUENCES ARE HIGHLY HOMOLOGOUS TO A REPEAT OF THE PDB3 MRNA OF DICTYOSTELIUM DISCOIDEUM.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: M16345; AAA33216.1; -			
DR	PIR: A26838; A26838.			
DR	HSSP: P00592; 3P2P.			
DR	DictyDb: DD03006; ecmb.			
DR	InterPro: IPR001673; S_mold_repeat.			
DR	Pfam: PF00526; S_mold_repeat; 42.			
DR	ProDom: PD006869; S_mold_repeat; 38.			
KW	Signal; Repeat.			
FT	SIGNAL 1 18			
FT	CHAIN 19 1046			
FT	DOMAIN 31 1012			
FT	REPEAT 31 53			
FT	REPEAT 34 77			
FT	REPEAT 54 101			
FT	REPEAT 102 125			
FT	REPEAT			
FT	A-1.			

PROBABLE.
PRESTALK PROTEIN.
41 X 24 AA TANDEM REPEATS. CYS-RICH.

FT	DOMAIN	26	64	LDL-RECEPTOR CLASS A 1.
FT	DOMAIN	65	105	LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	106	144	LDL-RECEPTOR CLASS A 3.
FT	DOMAIN	145	181	LDL-RECEPTOR CLASS A 4.
FT	DOMAIN	181	219	LDL-RECEPTOR CLASS A 5.
FT	DOMAIN	220	258	LDL-RECEPTOR CLASS A 6.
FT	DOMAIN	263	307	LDL-RECEPTOR CLASS A 7.
FT	DOMAIN	308	346	EGF-Like 1.
FT	DOMAIN	347	385	EGF-Like 2.
FT	DOMAIN	435	476	LDL-RECEPTOR CLASS B 1.
FT	DOMAIN	478	519	LDL-RECEPTOR CLASS B 2.
FT	DOMAIN	521	566	LDL-RECEPTOR CLASS B 3.
FT	DOMAIN	568	611	LDL-RECEPTOR CLASS B 4.
FT	DOMAIN	612	652	LDL-RECEPTOR CLASS B 5.
FT	DOMAIN	658	704	EGF-Like 3.
FT	DOMAIN	752	793	LDL-RECEPTOR CLASS B 6.
FT	DOMAIN	795	835	LDL-RECEPTOR CLASS B 7.
FT	DOMAIN	837	879	LDL-RECEPTOR CLASS B 8.
FT	DOMAIN	881	923	LDL-RECEPTOR CLASS B 9.
FT	DOMAIN	969	1013	EGF-Like 4.
FT	DOMAIN	1023	1061	LDL-RECEPTOR CLASS A 8.
FT	DOMAIN	1064	1103	LDL-RECEPTOR CLASS A 9.
FT	DOMAIN	1108	1146	LDL-RECEPTOR CLASS A 10.
FT	DOMAIN	1148	1186	LDL-RECEPTOR CLASS A 11.
FT	DOMAIN	1186	1225	LDL-RECEPTOR CLASS A 12.
FT	DOMAIN	1229	1269	LDL-RECEPTOR CLASS A 13.
FT	DOMAIN	1270	1308	LDL-RECEPTOR CLASS A 14.
FT	DOMAIN	1311	1351	LDL-RECEPTOR CLASS A 15.
FT	DOMAIN	1350	1390	EGF-Like 5.
FT	DOMAIN	1391	1430	EGF-Like 6, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1479	1520	LDL-RECEPTOR CLASS B 10.
FT	DOMAIN	1522	1563	LDL-RECEPTOR CLASS B 11.
FT	DOMAIN	1567	1609	LDL-RECEPTOR CLASS B 12.
FT	DOMAIN	1611	1654	LDL-RECEPTOR CLASS B 13.
FT	DOMAIN	1656	1696	LDL-RECEPTOR CLASS B 14.
FT	DOMAIN	1701	1742	EGF-Like 7.
FT	DOMAIN	1791	1832	LDL-RECEPTOR CLASS B 15.
FT	DOMAIN	1834	1882	LDL-RECEPTOR CLASS B 16.
FT	DOMAIN	1884	1930	LDL-RECEPTOR CLASS B 17.
FT	DOMAIN	1932	1972	LDL-RECEPTOR CLASS B 18.
FT	DOMAIN	1973	2013	LDL-RECEPTOR CLASS B 19.
FT	DOMAIN	2019	2060	EGF-Like 8.
FT	DOMAIN	2108	2156	LDL-RECEPTOR CLASS B 20.
FT	DOMAIN	2158	2201	LDL-RECEPTOR CLASS B 21.
FT	DOMAIN	2203	2245	LDL-RECEPTOR CLASS B 22.
FT	DOMAIN	2247	2289	LDL-RECEPTOR CLASS B 23.
FT	DOMAIN	2291	2332	LDL-RECEPTOR CLASS B 24.
FT	DOMAIN	2343	2384	EGF-Like 9.
FT	DOMAIN	2432	2477	LDL-RECEPTOR CLASS B 25.
FT	DOMAIN	2479	2518	LDL-RECEPTOR CLASS B 26.
FT	DOMAIN	2520	2562	LDL-RECEPTOR CLASS B 27.
FT	DOMAIN	2564	2604	LDL-RECEPTOR CLASS B 28.
FT	DOMAIN	2605	2647	LDL-RECEPTOR CLASS B 29.
FT	DOMAIN	2652	2694	EGF-Like 10.
FT	DOMAIN	2699	2739	LDL-RECEPTOR CLASS A 16.
FT	DOMAIN	2740	2778	LDL-RECEPTOR CLASS A 17.
FT	DOMAIN	2779	2820	LDL-RECEPTOR CLASS A 18.
FT	DOMAIN	2821	2862	LDL-RECEPTOR CLASS A 19.
FT	DOMAIN	2863	2903	LDL-RECEPTOR CLASS A 20.
FT	DOMAIN	2906	2947	LDL-RECEPTOR CLASS A 21.
FT	DOMAIN	2948	2992	LDL-RECEPTOR CLASS A 22.
FT	DOMAIN	2993	3031	LDL-RECEPTOR CLASS A 23.
FT	DOMAIN	3032	3072	LDL-RECEPTOR CLASS A 24.
FT	DOMAIN	3075	3112	LDL-RECEPTOR CLASS A 25.
FT	DOMAIN	3113	3153	EGF-Like 11.
FT	DOMAIN	3154	3194	EGF-Like 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	3241	3282	LDL-RECEPTOR CLASS B 30.
FT	DOMAIN	3284	3333	LDL-RECEPTOR CLASS B 31.
FT	DOMAIN	3335	3377	LDL-RECEPTOR CLASS B 32.
FT	DOMAIN	3379	3420	LDL-RECEPTOR CLASS B 33.
FT	DOMAIN	3421	3461	LDL-RECEPTOR CLASS B 34.
FT	DOMAIN	3467	3511	EGF-Like 13.
FT	DOMAIN	3512	3552	LDL-RECEPTOR CLASS A 26.

FT	DOMAIN	3553	3593	LDL-RECEPTOR CLASS A 27.
FT	DOMAIN	3594	3634	LDL-RECEPTOR CLASS A 28.
FT	DOMAIN	3635	3675	LDL-RECEPTOR CLASS A 29.
FT	DOMAIN	3678	3718	LDL-RECEPTOR CLASS A 30.
FT	DOMAIN	3719	3758	LDL-RECEPTOR CLASS A 31.
FT	DOMAIN	3759	3797	LDL-RECEPTOR CLASS A 32.
FT	DOMAIN	3798	3836	LDL-RECEPTOR CLASS A 33.
FT	DOMAIN	3842	3882	LDL-RECEPTOR CLASS A 34.
FT	DOMAIN	3883	3924	LDL-RECEPTOR CLASS A 35.
FT	DOMAIN	3928	3966	LDL-RECEPTOR CLASS A 36.
FT	DOMAIN	3968	4008	EGF-Like

RT	homolog of Drosophila Notch.";		
RT	Genomics 15:259-264(1993).		
RM	[2]		
RP	SEQUENCE OF 1551-2170 FROM N.A.		
RC	TISSUE-Embryo:		
RX	MEDLINE=93048835; PubMed=1425352;		
RA	Francisco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,		
RA	Greenspan R.J., McMahon A.P., Gridley T.;		
RT	"Expression pattern of Notch, a mouse homolog of Drosophila Notch,		
RT	suggests an important role in early postimplantation mouse		
RT	development.";		
RL	Development 115:737-744(1992).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.		
CC	-1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.		
CC	-1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.		
CC	-1- SIMILARITY: CONTAINS 5 ANK REPEATS.		
CC	-1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; Z11886; CAA77941.1; -.		
DR	HSSP; P00740; 1EDM.		
DR	MCD; MGI:97363; Notchl.		
DR	InterPro; IPR002110; ANK.		
DR	InterPro; IPR000152; Asx_hydroxyl.		
DR	InterPro; IPR000561; EGF_Like.		
DR	InterPro; IPR000742; EGF_2.		
DR	InterPro; IPR001881; EGF_Ca.		
DR	InterPro; IPR001438; EGF_II.		
DR	InterPro; IPR000800; Notch.		
DR	Pfam; PF00023; ank; 6.		
DR	Pfam; PF00008; EGF; 35.		
DR	Pfam; PF00066; notch; 3.		
DR	PRINTS; PR00010; EGFBLD.		
DR	PRINTS; PR01452; NOTCH.		
DR	SMART; SM00248; ANK_3.		
DR	SMART; SM00179; EGF_CA; 23.		
DR	SMART; SM00001; EGF_Like; 11.		
DR	SMART; SM00004; NL; 2.		
DR	PROSITE; PS50086; ANK_REPEAT; 2.		
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.		
DR	PROSITE; PS00010; ASX_HYDROXYL; 22.		
DR	PROSITE; PS00022; EGF_1; 34.		
DR	PROSITE; PS01186; EGF_2; 27.		
DR	PROSITE; PS01187; EGF_CA; 21.		
KW	Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;		
KW	Transmembrane; Signal; Glycoprotein.		
FT	SIGNAL 1 18		
FT	CHAIN 19 2531		
FT	DOMAIN 19 1725		
FT	TRANSMEM 1726 1746		
FT	DOMAIN 1747 2531		
FT	DOMAIN 20 58		
FT	DOMAIN 59 99		
FT	DOMAIN 102 139		
FT	DOMAIN 140 176		
FT	DOMAIN 178 216		
FT	DOMAIN 181 255		
FT	DOMAIN 257 293		
FT	DOMAIN 295 333		
FT	DOMAIN 335 371		
FT	DOMAIN 372 410		
FT	DOMAIN 412 450		
FT	DOMAIN 452 488		
FT	DOMAIN 490 526		
FT	DOMAIN 528 564		
FT	DOMAIN 564		

FT	DOMAIN	566	601	EGF-LIKE 15,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	603	639	EGF-LIKE 16,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	641	676	EGF-LIKE 17,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	678	714	EGF-LIKE 18,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	716	751	EGF-LIKE 19,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	753	789	EGF-LIKE 20,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	791	827	EGF-LIKE 21,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	829	867	EGF-LIKE 22,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	869	905	EGF-LIKE 23,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	907	943	EGF-LIKE 24,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	945	981	EGF-LIKE 25,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	983	1019	EGF-LIKE 26,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1021	1057	EGF-LIKE 27,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1059	1095	EGF-LIKE 28,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1097	1143	EGF-LIKE 29,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1145	1181	EGF-LIKE 30,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1183	1219	EGF-LIKE 31,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1221	1265	EGF-LIKE 32,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1267	1305	EGF-LIKE 33,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1307	1346	EGF-LIKE 34,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1387	1426	EGF-LIKE 35,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	1449	1462	CYS-RICH,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	1445	1480	LIN/NOTCH 1,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	1481	1522	LIN/NOTCH 2,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	1523	1562	LIN/NOTCH 3,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	1917	1947	ANK 1,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	1917	1979	ANK 2,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	1983	2012	ANK 3,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	2016	2045	ANK 4,	CALCIUM-BINDING	(POTENTIAL)
FT	REPEAT	2049	2078	ANK 5,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	24	37	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	31	46	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	63	74	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	68	87	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	89	98	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	106	117	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	111	127	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	129	138	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	144	155	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	149	164	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	166	175	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	182	195	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	189	204	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	206	215	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	222	233	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	227	243	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	245	254	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	261	272	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	266	281	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	283	292	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	299	312	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	306	321	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	323	332	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	339	350	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	344	359	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	361	370	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	376	387	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	381	398	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	400	409	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	416	429	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	423	438	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	440	449	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	456	467	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	461	476	BY SIMILARITY,	CALCIUM-BINDING	(POTENTIAL)
FT	DISULFID	4				

```

FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 987 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.

```

Query Match Best Local Similarity 17.7%; Score 117.5; DB 1; Length 2531; Pred. No. 0.00411; Matches 35; Conservative 7; Mismatches 46; Indels 21; Gaps 6;

```

QY 15 IAVIWC-----LCFE-----CYCHCTGAGDCSTCTACTGCG--GNCNPAHTCTD--SKNC 61
DB 1110 IIVTLICGQGLGCTVDEBGRKHCHGAGTGYCCEDEVEBCSPNQGNATCTDYIGGRSC 1169
QY 62 VAAATCTGCTKCNARTCTNSKDCFEAKTCTDSTNCYATACTNSTGCP 110
DB 11170 KCVAGYHGSCSEINECL-SQPCONGGTCTDITNSYKSC-----CP 1210

```

```

RESULT 4
NTCL_RAT STANDARD: PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuronic locus notch homolog protein 1 precursor.
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
RA Weinmaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian
RT development."
RL Development 113:199-205(1991).
RN 12
RP REVISIONS TO 1652-1653.

```

```

RA Weinmaster G.;
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
CC OF TISSUES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
CC ADULT.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57405; CAA40667.1; -.
DR HSSP; P00740; IIXA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00011; EGFFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 25.
DR SMART; SM00001; EGF_Like; 10.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 35.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS01187; EGF_CA; 21.
DR Transmembrane; Signal; Glycoprotein.
KW SIGNAL.
FT CHAIN 1 18
FT DOMAIN 19 2531
FT TRANSMEM 1724 1746
FT DOMAIN 1747 2531
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751

```

```

FT DOMAIN 753 789 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 867 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 907 943 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 945 981 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 983 1019 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1021 1057 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1059 1095 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1097 1143 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1267 1305 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1307 1346 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1348 1384 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1387 1426 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1449 1462 CYS-RICH.
FT REPEAT 1917 1946 ANK 1.
FT REPEAT 1950 1980 ANK 2.
FT REPEAT 1984 2013 ANK 3.
FT REPEAT 2017 2046 ANK 4.
FT REPEAT 2050 2079 ANK 5.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.

```

```

FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 747 757 BY SIMILARITY.
FT DISULFID 752 768 BY SIMILARITY.
FT DISULFID 759 777 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 845 BY SIMILARITY.
FT DISULFID 857 866 BY SIMILARITY.
FT DISULFID 873 884 BY SIMILARITY.
FT DISULFID 878 893 BY SIMILARITY.
FT DISULFID 895 904 BY SIMILARITY.
FT DISULFID 911 922 BY SIMILARITY.
FT DISULFID 916 931 BY SIMILARITY.
FT DISULFID 933 942 BY SIMILARITY.
FT DISULFID 937 948 BY SIMILARITY.
FT DISULFID 967 998 BY SIMILARITY.
FT DISULFID 992 1007 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1025 1036 BY SIMILARITY.
FT DISULFID 1030 1045 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1063 1074 BY SIMILARITY.
FT DISULFID 1068 1083 BY SIMILARITY.
FT DISULFID 1085 1094 BY SIMILARITY.
FT DISULFID 1101 1122 BY SIMILARITY.
FT DISULFID 1116 1131 BY SIMILARITY.

```

Query Match 17.7% Score 117.5; DB 1; Length 2531;
 Best Local Similarity 32.1%; Pred. No. 0.0041;
 Matches 35; Conservative 7; Mismatches 46; Indels 21; Gaps 6;

```

QY 15 IAVIWC-----LCTE-----CYCHCTGACDCTCTACTGCG--GNCPRNHTCTD---SKNC 61
Db 1110 IDVTLLCQHGGLCVDEEDKHCHCQAGTIGSTCEDEVCSPNFCQNKATCTDIYIGRSC 1169
QY 62 VKAATCTGTFKCNTRATCTNSKDCFEAKTCTDSTNCRATACSTNGSCP 110
Db 1170 KCVAGYHSGNCSESEINECL-SQPCQNGGTCTDITWTYCS-----CP 1210

```

```

RESULT 5
KRUB_HUMAN STANDARD; PRT; 194 AA.
AC 075690;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, ultra high-sulfur matrix protein B (UHS keratin B) (UHS
  DE Kerb).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148005; Pubmed=10023043;
RA Perez C., Aurio J., Gerst C., Bernard B.A., Egly J.-M.;
RT "Genomic organization and promoter characterization of two human UHS
  RT keratin genes.";
RL Gene 227:137-148(1999).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
  CC SUCH AS WOOL AND HAIR CONSIST OF MICROFILBRILS EMBEDDED IN A RIGID
  CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
  CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
  CC 6-20 kDa, WHEREAS THE MICROFILBRILS CONTAIN THE LARGER, LOW-SULFUR
  CC KERATINS (40-56 kDa).

```



```

FT REPEAT 1928 1957 ANK 1.
FT REPEAT 1961 1991 ANK 2.
FT REPEAT 1995 2024 ANK 3.
FT REPEAT 2028 2057 ANK 4.
FT REPEAT 2061 2090 ANK 5.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1662 1665 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 589 599 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.

```

```

FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 867 BY SIMILARITY.
FT DISULFID 874 885 BY SIMILARITY.
FT DISULFID 879 894 BY SIMILARITY.
FT DISULFID 896 905 BY SIMILARITY.
FT DISULFID 912 923 BY SIMILARITY.
FT DISULFID 917 932 BY SIMILARITY.
FT DISULFID 934 943 BY SIMILARITY.
FT DISULFID 988 999 BY SIMILARITY.
FT DISULFID 993 1008 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1026 1037 BY SIMILARITY.
FT DISULFID 1031 1046 BY SIMILARITY.
FT DISULFID 1048 1057 BY SIMILARITY.
FT DISULFID 1064 1075 BY SIMILARITY.

```

Query Match 17.2%; Score 114; DB 1; Length 2444;

Best Local Similarity 31.6%; Pred. No. 0.0081; Matches 31; Conservative 6; Mismatches 45; Indels 16; Gaps 6;

```

QY 23 CTE----CYCHCTGADCTCTDCTGCGN--CPNAHCTDSKNCVKATCT-----GST 71
Db 923 CTGGINTPAFCDLPGRGRGFCEEDINECASDPCRNANCND---CYDSTYTCIPAFSGI 979
QY 72 KC-NTARTCTNSKDCFEAKTCTDSTNCYKATACTNSTG 108
Db 980 HCENNTPDCES--SCFNGTCVGDGINSFTCLCPPGFTG 1016

```

RESULT 7

```

ID FBNI_HUMAN STANDARD: PRT; 2871 AA.
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 1 precursor.
GN FBNI OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Pangillan T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome.";
RL Hum. Mol. Genet. 2:961-968(1993).
RN [2]
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta, and fibroblast;
RX MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
RT end.";
RL Genomics 17:476-484(1993).
RN [3]
RP SEQUENCE OF 899-2871 FROM N.A.
RX MEDLINE=91304568; PubMed=1852207;
RA Maslen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT "Partial sequence of a candidate gene for the Marfan syndrome.";
RL Nature 352:334-337(1991).
RN [4]
RP SEQUENCE OF 813-1313 FROM N.A.
RX MEDLINE=91304567; PubMed=1852206;

```


RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA Tsiipouras P., Ramirez F., Hollister D.W.;
RA "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes."; *Nature* 352:330-334(1991).
RN [15]
RN CHARACTERIZATION.
RX MEDLINE=91317849; PubMed=1860873;
RA Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;
RT "Purification and partial characterization of fibrillin, a cysteine-
RT rich structural component of connective tissue microfibrils."; *J.
RL Biol. Chem.* 266:14763-14770(1991).
RN [16]
RN STRUCTURE BY NMR OF 2054-2125.
RX MEDLINE=98031893; PubMed=9362480;
RA Yuan X., Downing A.K., Knott V., Handford P.A.;
RT "Solution structure of the transforming growth factor beta-binding
RT protein-like module, a domain associated with matrix fibrils."; *EMBO J.*
RL 16:6659-6666(1997).
RN [17]
RN STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE=96144829; PubMed=856869;
RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;
RT "Calcium binding properties of an epidermal growth factor-like domain
RT pair from human fibrillin-1."; *J. Mol. Biol.* 255:22-27(1996).
RN [18]
RN STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE=9622301; PubMed=8653794;
RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,
RA Handford P.A.;
RT "Solution structure of a pair of calcium-binding epidermal growth
RT factor-like domains: implications for the Marfan syndrome and other
RT genetic disorders."; *Cell* 85:597-605(1996).
RN [19]
RN REVIEW ON MFS VARIANTS.
RX MEDLINE=96174615; PubMed=8594563;
RA Colod-C, Beroud C., Soussi T., Junten C., Bollenau C.;
RT "Software and database for the analysis of mutations in the human
RT FBN1 gene."; *Nucleic Acids Res.* 24:137-141(1996).
RN [10]
RN REVIEW ON MFS VARIANTS.
RX MEDLINE=97169383; PubMed=9016526;
RA Colod-C, Beroud C., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,
RA Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,
RA Richards R.I., Wang W., Junten C., Bollenau C.;
RT "Marfan Database (second edition): software and database for the
RT analysis of mutations in the human FBN1 gene."; *Nucleic Acids Res.*
RL 25:147-150(1997).
RN [11]
RN NUCLEIC ACIDS RES. 25:147-150(1997).
RN REVIEW ON VARIANTS.
RX MEDLINE=98062175; PubMed=9401003;
RA Hayward C., Brock D.J.H.;
RT "Fibrillin-1 mutations in Marfan syndrome and other type-1
RT fibrillinopathies."; *Hum. Mutat.* 10:415-423(1997).
RN [12]
RN VARIANT MFS PRO-1137.
RX MEDLINE=91304569; PubMed=1852208;
RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
RA Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;
RT "Marfan syndrome caused by a recurrent de novo missense mutation in
RT the fibrillin gene."; *Nature* 352:337-339(1991).
RN [13]
RN VARIANT MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.
RX MEDLINE=93250834; PubMed=1301946;
RA Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
RT patients at cysteine residues in EGF-like domains."; *Hum. Mutat.* 1:366-374(1992).
RN [14]
RN VARIANT MFS SER-2307.
RX MEDLINE=92235290; PubMed=1569206;
RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,
RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
RT "Marfan phenotype variability in a family segregating a missense
RT mutation in the epidermal growth factor-like motif of the fibrillin
RT gene."; *J. Clin. Invest.* 89:1674-1680(1992).
RN [15]
RN VARIANTS MFS ILE-548 AND ALA-723.
RX MEDLINE=94010946; PubMed=8406497;
RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
RA Pyeritz R.E., Francomano C.A.;
RT "Four novel FBN1 mutations: significance for mutant transcript level
RT and EGF-like domain calcium binding in the pathogenesis of Marfan
RT syndrome."; *Genomics* 17:468-475(1993).
RN [16]
RN VARIANTS MFS SER-2144.
RX MEDLINE=93278402; PubMed=8504310;
RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
RT "A novel fibrillin mutation in the Marfan syndrome which could
RT disrupt calcium binding of the epidermal growth factor-like module."; *Hum. Mol. Genet.* 2:475-477(1993).
RN [17]
RN VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.
RX MEDLINE=94108431; PubMed=8281141;
RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
RA Berg M.A., Miller D.C., Francke U.;
RT "Mutation screening of complete fibrillin-1 coding sequence: report
RT of five new mutations, including two in 8-cysteine domains."; *Hum. Mol. Genet.* 2:1813-1821(1993).
RN [18]
RN VARIANTS MFS GLY-217 AND ARG-2627.
RX MEDLINE=95067970; PubMed=7977366;
RA Karttunen L., Raghunath M., Loengqvist L., Peltonen L.;
RT "A compound heterozygous Marfan patient: two defective fibrillin
RT alleles result in a lethal phenotype."; *Am. J. Hum. Genet.* 55:1083-1091(1994).
RN [19]
RN VARIANT EL LYS-2447.
RX MEDLINE=94245249; PubMed=8188302;
RA Longqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
RA Peltonen L.;
RT "A novel mutation of the fibrillin gene causing ectopia lentis."; *Genomics* 19:573-576(1994).
RN [20]
RN VARIANT MFS CYS-627.
RX MEDLINE=94272487; PubMed=8004112;
RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
RT "Two novel mutations and a neutral polymorphism in EGF-like domains
RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
RT syndrome patients."; *Hum. Mol. Genet.* 3:373-375(1994).
RN [21]
RN VARIANT MFS CYS-122.
RX MEDLINE=94314977; PubMed=8040326;
RA Stahl-Hallengren C., Ukonen T., Kainulainen K., Kristofersson U.,
RA Saxne T., Tornqvist K., Peltonen L.;
RT "An extra cysteine in one of the non-calcium-binding epidermal growth
RT factor-like motifs of the FBN1 polypeptide is connected to a novel
RT variant of Marfan syndrome."; *J. Clin. Invest.* 94:709-713(1994).
RN [22]
RN VARIANT MFS TYR-1223.
RX MEDLINE=94351682; PubMed=8071963;
RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
RT "A new missense mutation of fibrillin in a patient with Marfan
RT syndrome."; *J. Med. Genet.* 31:338-339(1994).
RN [23]

```

RP VARIANT MFS HIS-1170.
RX MEDLINE=95174777; PubMed=7870075.
RA Hayward C., Porteous M.E.M., Brock D.J.H.;
RT "A novel mutation in the fibrillin gene (FBN1) in familial
RL arachnodactyly.";
RN Mol. Cell. Probes 8:325-327(1994).
RP [24]
RP VARIANTS MFS G-217; N-1023; R-1074; Y-1242; R-1513; E-2127; W-2151;
RX K-2447 AND R-2511.
RP MEDLINE=94184368; PubMed=8136837;
RA Kainulainen K., Karttunen L., Punakka L., Sakai L., Peltomen L.;
RT "Mutations in the fibrillin gene responsible for dominant ectopia
RL lentis and neonatal Marfan syndrome.";
RN Nat. Genet. 6:64-69(1994).
RP [25]
RP VARIANT SER-1127.

Query Match 17.1%; Score 113.5; DB 1; Length 2871;
Best Local Similarity 27.1%; Pred. No. 0.01;
Matches 36; Conservative 9; Mismatches 41; Indels 47; Gaps 7;

QY 23 CT-----ECYCH-----CTGAGDCTCTDACTG--CGNCNATTC----- 55
DB 1212 CTNBSGSEYECSCQPFALMPDQRCSTCIDCEDNPNICDGGQCTNIGEXRCICYDGFMA 1271
QY 56 -TDSKNCVKATCTGTFKCTARTCTNSKDCF-----EAKT-CTD-----ST 95
DB 1272 SEDMKTCVDVMECDLNPNICISGTCENTKGSFICHCDMGYSGKRGKTCGTDINECEIGAH 1331
QY 96 NCYKATACSTNG 108
DB 1332 NCGKHAVCTNTAG 1344

RESULT 8
NOTCH_XENLA STANDARD; PRU; 2524 AA.
ID NOTCH_XENLA
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update).
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein).
OS XOTCH.
GN Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RP Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1 SUBCELLULAR LOCATION: Type I membrane protein.
CC -1 DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1 SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1 SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1 SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1 SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: M3874; AAB02039.1;
DR PIR: A35844; A35844.
DR HSPR: P00740; 1EDM.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-1like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001438; EGF-IT.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF_CA; 23.
DR Pfam: PF00006; EGF; 36.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFLOOD.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_CA; 23.
DR SMART: SM00001; EGF_1like; 11.
DR SMART: SM00004; NL; 2.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 23.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01186; EGF_2; 29.
DR PROSITE: PS01187; EGF_CA; 21.
DR KEGG: Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 2524
FT DOMAIN 20 1728 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT TRANSMEM 1729 1750 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1751 2524 POTENTIAL.
FT DOMAIN 20 57 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 99 EGF-LIKE 1.
FT DOMAIN 102 140 EGF-LIKE 2.
FT DOMAIN 141 177 EGF-LIKE 3.
FT DOMAIN 179 215 EGF-LIKE 4.
FT DOMAIN 217 254 EGF-LIKE 5.
FT DOMAIN 256 292 EGF-LIKE 6.
FT DOMAIN 294 332 EGF-LIKE 7.
FT DOMAIN 334 370 EGF-LIKE 8.
FT DOMAIN 371 409 EGF-LIKE 9.
FT DOMAIN 411 449 EGF-LIKE 10.
FT DOMAIN 451 487 EGF-LIKE 11.
FT DOMAIN 489 525 EGF-LIKE 12.
FT DOMAIN 527 563 EGF-LIKE 13.
FT DOMAIN 565 600 EGF-LIKE 14.
FT DOMAIN 602 638 EGF-LIKE 15.
FT DOMAIN 640 675 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 750 EGF-LIKE 18.
FT DOMAIN 752 788 EGF-LIKE 19.
FT DOMAIN 790 826 EGF-LIKE 20.
FT DOMAIN 828 866 EGF-LIKE 21.
FT DOMAIN 868 904 EGF-LIKE 22.
FT DOMAIN 906 942 EGF-LIKE 23.
FT DOMAIN 944 980 EGF-LIKE 24.
FT DOMAIN 982 1018 EGF-LIKE 25.
FT DOMAIN 1020 1056 EGF-LIKE 26.
FT DOMAIN 1058 1094 EGF-LIKE 27.
FT DOMAIN 1096 1142 EGF-LIKE 28.
FT DOMAIN 1144 1180 EGF-LIKE 29.
FT DOMAIN 1182 1218 EGF-LIKE 30.
FT DOMAIN 1220 1264 EGF-LIKE 31.
FT DOMAIN 1266 1304 EGF-LIKE 32.
FT DOMAIN 1306 1346 EGF-LIKE 33.
FT DOMAIN 1347 1383 EGF-LIKE 34.
FT DOMAIN 1386 1424 EGF-LIKE 35.
FT DOMAIN 1441 1478 EGF-LIKE 36.
FT REPEAT 1479 1520 LIN/NOTCH 1.
FT REPEAT 1521 1560 LIN/NOTCH 2.
FT REPEAT 1561 1600 LIN/NOTCH 3.
FT REPEAT 1601 1640 ANK 1.
FT REPEAT 1641 1680 ANK 2.
FT REPEAT 1681 1720 ANK 3.
FT REPEAT 1721 1760 ANK 4.
FT REPEAT 1761 1800 ANK 5.
FT REPEAT 1801 1840 ANK 6.
FT REPEAT 1841 1880 ANK 7.
FT REPEAT 1881 1920 ANK 8.
FT REPEAT 1921 1960 ANK 9.
FT REPEAT 1961 2000 ANK 10.
FT REPEAT 2001 2040 ANK 11.
FT REPEAT 2041 2080 ANK 12.
FT REPEAT 2081 2120 ANK 13.
FT REPEAT 2121 2160 ANK 14.
FT REPEAT 2161 2200 ANK 15.
FT REPEAT 2201 2240 ANK 16.
FT REPEAT 2241 2280 ANK 17.
FT REPEAT 2281 2320 ANK 18.
FT REPEAT 2321 2360 ANK 19.
FT REPEAT 2361 2400 ANK 20.
FT REPEAT 2401 2440 ANK 21.
FT REPEAT 2441 2480 ANK 22.
FT REPEAT 2481 2524 ANK 23.

```

FT	REPEAT	1957	1987	ANK 3.
FT	REPEAT	1991	2020	ANK 4.
FT	REPEAT	2024	2053	ANK 5.
FT	REPEAT	2057	2086	ANK 6.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	504	BY SIMILARITY.
FT	DISULFID	498	513	BY SIMILARITY.
FT	DISULFID	515	524	BY SIMILARITY.
FT	DISULFID	531	542	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	553	562	BY SIMILARITY.
FT	DISULFID	569	579	BY SIMILARITY.
FT	DISULFID	574	588	BY SIMILARITY.
FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	617	BY SIMILARITY.
FT	DISULFID	611	626	BY SIMILARITY.
FT	DISULFID	628	637	BY SIMILARITY.
FT	DISULFID	644	654	BY SIMILARITY.
FT	DISULFID	649	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	729	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	756	767	BY SIMILARITY.
FT	DISULFID	761	776	BY SIMILARITY.
FT	DISULFID	778	787	BY SIMILARITY.
FT	DISULFID	794	805	BY SIMILARITY.
FT	DISULFID	799	814	BY SIMILARITY.
FT	DISULFID	816	825	BY SIMILARITY.
FT	DISULFID	832	843	BY SIMILARITY.
FT	DISULFID	837	854	BY SIMILARITY.
FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.

FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	966	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.
FT	DISULFID	1132	1141	BY SIMILARITY.
FT	DISULFID	1148	1159	BY SIMILARITY.
FT	DISULFID	1153	1168	BY SIMILARITY.
Query Match 17.0%; Score 113; DB 1; Length 2524;				
Best Local Similarity 34.3%; Pred. No. 0.01;				
Matches 34; Conservative 5; Mismatches 42; Indels 18; Gaps 6;				
QY	23	CTECY----CHCT---GGADCTSCTDACGCGNCPNAAHTCTDSKNCVKAATC-----TG 69		
Db	959	CTDCVNSYTCQCPGSGHICESNFPDCTE--SSCFNGGTCIDG---INTFTGCPGFTG 1014		
QY	70	STKCNARTCTNKKDFEAKTCTDSTNCKAKATACINSTG 108		
Db	1015	STCQHDINEC-DSKPLNGCTGDSYGYKCYCPQGYTG 1052		
RESULT 9				
ID	KRUA_HUMAN	STANDARD;	PRT;	169 AA.
AC	P26371;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Keratin, ultra high-sulfur matrix protein A (UHS keratin A) (UHS			
DE	kerat).			
GN	KRNI OR UHSK1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=follicle;			
RX	MEDLINE=91115951; PubMed=1703541;			
RA	McKinnon P.J., Powell B.C., Rogers G.E.;			
RT	"Structure and expression of genes for a class of cysteine-rich			
RT	proteins of the cuticle layers of differentiating wool and hair			
RT	follicles.";			
RL	J. Cell Biol. 111:2587-2600(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99148005; PubMed=10023043;			
RA	Perez C., Auriool J., Gerst C., Bernard B.A., Egly J.-M.;			
RT	"Genomic organization and promoter characterization of two human UHS			
RT	keratin genes.";			
RL	Gene 227:137-148(1999).			
FC	-1- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES			
CC	SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID			
CC	MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-			
CC	SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF			
CC	6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR			
CC	KERATINS (40-56 kDa).			
CC	-1- TISSUE SPECIFICITY: CUTICLE LAYERS OF DIFFERENTIATING			
CC	HAIR FOLLICLES.			
CC	-1- DEVELOPMENTAL STAGE: AT A LATE STAGE OF FIBER DIFFERENTIATION.			
CC	-1- DOMAIN: MAINLY COMPOSED OF CYS-RICH (CR), GLY-RICH (GR) AND SER-			
CC	RICH (SR) REPEATS.			
CC	-1- SIMILARITY: BELONGS TO THE UHS KERATIN FAMILY.			

FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING.
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING.
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING.
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING.
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING.
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING.
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.
FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.
FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1166	1180	BY SIMILARITY.
FT	DISULFID	1182	1195	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1238	1254	BY SIMILARITY.
FT	DISULFID	1242	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.
FT	DISULFID	1326	1339	BY SIMILARITY.
FT	DISULFID	1333	1348	BY SIMILARITY.
FT	DISULFID	1350	1361	BY SIMILARITY.
FT	DISULFID	1367	1380	BY SIMILARITY.

FT	DISULFID	1374	1389	BY SIMILARITY.
FT	DISULFID	1391	1402	BY SIMILARITY.
FT	DISULFID	1408	1420	BY SIMILARITY.
FT	DISULFID	1415	1429	BY SIMILARITY.
FT	DISULFID	1431	1444	BY SIMILARITY.
FT	DISULFID	1450	1461	BY SIMILARITY.
FT	DISULFID	1456	1470	BY SIMILARITY.
FT	DISULFID	1472	1485	BY SIMILARITY.
FT	DISULFID	1491	1502	BY SIMILARITY.
FT	DISULFID	1497	1511	BY SIMILARITY.
FT	DISULFID	1513	1526	BY SIMILARITY.
FT	DISULFID	1610	1622	BY SIMILARITY.

Query Match 16.6%; Score 110.5; DB 1; Length 2871;
 Best Local Similarity 26.3%; Pred. No. 0.018;
 Matches 35; Conservative 10; Mismatches 41; Indels 47; Gaps 7;

QY	23	CT-----ECYCH-----	CTGADCTSCCTDAGT--CGNCPNAHTC-----	55
DB	1212	CTNSEGSYCSGSGPFLMPDQRCSTHDECDNPNICGGGCTNIPGEYRCLCYDGFMA	1271	
QY	56	-TDSKNCVAAATCTGCTKCTARTCTNSKDCF-----	EAKT-CTD-----	ST 95
DB	1272	SEDMKTCVVDNEDLNPNICLSGTCEMTKGSFTCHODMGSGKKGTGCTDINECEIGAH	1331	
QY	96	NCYKATACNTSTG	108	
DB	1332	NCDRAVCTNTAG	1344	

RESULT 11
 FBNI_PIG
 ID FBNI_PIG STANDARD; PRT: 2871 AA.
 AC Q9TV36;
 DT 16-OCT-2001 (rel. 40; Created)
 DT 16-OCT-2001 (rel. 40; Last sequence update)
 DE Fibrillin 1 precursor.
 GN FBNI.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99156858; PubMed=10036187;
 RA Biery N.J.; Eldadah Z.A.; Moore C.S.; Stetten G.; Spencer F.;
 RA Dietz H.C.;
 RT "Revised genomic organization of FBNI and significance for regulated
 gene expression".
 RL Genomics 56:70-77(1999).
 CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
 THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
 LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
 FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
 MICROFIBRILS.
 CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF073800; AAD50328.1; -.
 CC HSP: P35555; IAPJ.
 CC InterPro: IPR000152; Asx_hydroxyl.

DR	InterPro: IPR000561; EGF-like.
DR	InterPro: IPR001881; EGF_Ca.
DR	InterPro: IPR001438; EGF_11.
DR	InterPro: IPR002212; TB.
DR	Pfam: PF00008; EGF; 45.
DR	Pfam: PF00683; TB; 9.
DR	PRINTS: PRO0010; EGFBL00D.
DR	SMART: SM00179; EGF_CA; 40.
DR	SMART: SM00001; EGF_Like; 6.
DR	PROSITE: PS00010; ASX_HYDROXYL; 41.
DR	PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR	PROSITE: PS01186; EGF_2; 36.
DR	PROSITE: PS01187; EGF_CA; 41.
KW	Extracellular matrix; Calcium binding; Glycoprotein; EGF-like domain.
KW	Repeat; Signal; Multigene family.
FT	SIGNAL 1 27
FT	CHAIN 28 2871
FT	DOMAIN 81 112
FT	DOMAIN 115 146
FT	DOMAIN 147 178
FT	DOMAIN 246 287
FT	DOMAIN 288 329
FT	REPEAT 330 390
FT	DOMAIN 392 446
FT	DOMAIN 449 489
FT	DOMAIN 490 529
FT	DOMAIN 530 571
FT	DOMAIN 572 612
FT	DOMAIN 613 653
FT	REPEAT 654 722
FT	DOMAIN 723 764
FT	DOMAIN 765 806
FT	DOMAIN 807 846
FT	DOMAIN 910 951
FT	REPEAT 952 1027
FT	DOMAIN 1028 1069
FT	DOMAIN 1070 1112
FT	DOMAIN 1113 1154
FT	DOMAIN 1155 1196
FT	DOMAIN 1197 1237
FT	DOMAIN 1238 1279
FT	DOMAIN 1280 1321
FT	DOMAIN 1322 1362
FT	DOMAIN 1363 1403
FT	DOMAIN 1404 1445
FT	DOMAIN 1446 1486
FT	DOMAIN 1487 1527
FT	REPEAT 1528 1605
FT	DOMAIN 1606 1647
FT	DOMAIN 1648 1688
FT	REPEAT 1689 1765
FT	DOMAIN 1766 1807
FT	DOMAIN 1808 1848
FT	DOMAIN 1849 1890
FT	DOMAIN 1891 1929
FT	DOMAIN 1930 1972
FT	DOMAIN 1973 2012
FT	DOMAIN 2013 2054
FT	REPEAT 2055 2126
FT	DOMAIN 2127 2165
FT	DOMAIN 2166 2205
FT	DOMAIN 2206 2246
FT	DOMAIN 2247 2290
FT	DOMAIN 2291 2332
FT	REPEAT 2333 2401
FT	DOMAIN 2402 2443
FT	DOMAIN 2444 2484
FT	DOMAIN 2485 2523
FT	DOMAIN 2524 2566
FT	DOMAIN 2567 2606
FT	DOMAIN 2607 2647
FT	DOMAIN 2648 2687
FT	DISULFID 85 94

FT	DISUFLID	89	100	BY SIMILARITY
FT	DISUFLID	102	111	BY SIMILARITY
FT	DISUFLID	119	129	BY SIMILARITY
FT	DISUFLID	123	134	BY SIMILARITY
FT	DISUFLID	136	145	BY SIMILARITY
FT	DISUFLID	150	160	BY SIMILARITY
FT	DISUFLID	154	166	BY SIMILARITY
FT	DISUFLID	168	177	BY SIMILARITY
FT	DISUFLID	250	261	BY SIMILARITY
FT	DISUFLID	257	271	BY SIMILARITY
FT	DISUFLID	273	286	BY SIMILARITY
FT	DISUFLID	292	304	BY SIMILARITY
FT	DISUFLID	299	313	BY SIMILARITY
FT	DISUFLID	315	328	BY SIMILARITY
FT	DISUFLID	435	465	BY SIMILARITY
FT	DISUFLID	460	474	BY SIMILARITY
FT	DISUFLID	476	488	BY SIMILARITY
FT	DISUFLID	494	504	BY SIMILARITY
FT	DISUFLID	499	513	BY SIMILARITY
FT	DISUFLID	515	528	BY SIMILARITY
FT	DISUFLID	534	546	BY SIMILARITY
FT	DISUFLID	541	556	BY SIMILARITY
FT	DISUFLID	557	570	BY SIMILARITY
FT	DISUFLID	576	587	BY SIMILARITY
FT	DISUFLID	582	596	BY SIMILARITY
FT	DISUFLID	598	611	BY SIMILARITY
FT	DISUFLID	617	628	BY SIMILARITY
FT	DISUFLID	623	637	BY SIMILARITY
FT	DISUFLID	639	652	BY SIMILARITY
FT	DISUFLID	727	739	BY SIMILARITY
FT	DISUFLID	734	748	BY SIMILARITY
FT	DISUFLID	750	763	BY SIMILARITY
FT	DISUFLID	769	781	BY SIMILARITY
FT	DISUFLID	776	790	BY SIMILARITY
FT	DISUFLID	792	805	BY SIMILARITY
FT	DISUFLID	811	821	BY SIMILARITY
FT	DISUFLID	816	830	BY SIMILARITY
FT	DISUFLID	832	845	BY SIMILARITY
FT	DISUFLID	914	926	BY SIMILARITY
FT	DISUFLID	921	935	BY SIMILARITY
FT	DISUFLID	937	950	BY SIMILARITY
FT	DISUFLID	1032	1044	BY SIMILARITY
FT	DISUFLID	1039	1053	BY SIMILARITY
FT	DISUFLID	1055	1068	BY SIMILARITY
FT	DISUFLID	1074	1086	BY SIMILARITY
FT	DISUFLID	1081	1095	BY SIMILARITY
FT	DISUFLID	1097	1111	BY SIMILARITY
FT	DISUFLID	1117	1129	BY SIMILARITY
FT	DISUFLID	1124	1138	BY SIMILARITY
FT	DISUFLID	1140	1153	BY SIMILARITY
FT	DISUFLID	1159	1171	BY SIMILARITY
FT	DISUFLID	1201	1212	BY SIMILARITY
FT	DISUFLID	1208	1221	BY SIMILARITY
FT	DISUFLID	1223	1236	BY SIMILARITY
FT	DISUFLID	1242	1254	BY SIMILARITY
FT	DISUFLID	1249	1263	BY SIMILARITY
FT	DISUFLID	1265	1278	BY SIMILARITY
FT	DISUFLID	1284	1296	BY SIMILARITY
FT	DISUFLID	1291	1305	BY SIMILARITY
FT	DISUFLID	1307	1320	BY SIMILARITY
FT	DISUFLID	1326	1339	BY SIMILARITY
FT	DISUFLID	1333	1348	BY SIMILARITY
FT	DISUFLID	1350	1361	BY SIMILARITY
FT	DISUFLID	1367	1380	BY SIMILARITY
FT	DISUFLID	1374	1389	BY SIMILARITY
FT	DISUFLID	1391	1402	BY SIMILARITY
FT	DISUFLID	1408	1420	BY SIMILARITY
FT	DISUFLID	1415	1429	BY SIMILARITY
FT	DISUFLID	1450	1461	BY SIMILARITY
FT	DISUFLID	1456	1470	BY SIMILARITY
FT	DISUFLID	1472	1485	BY SIMILARITY
FT	DISUFLID	1491	1502	BY SIMILARITY
FT	DISUFLID	1497	1511	BY SIMILARITY

FT DISULFID 1513 1526 BY SIMILARITY.
 FT DISULFID 1610 1622 BY SIMILARITY.
 FT DISULFID 1617 1631 BY SIMILARITY.
 FT DISULFID 1633 1646 BY SIMILARITY.
 FT DISULFID 1652 1663 BY SIMILARITY.
 FT DISULFID 1658 1672 BY SIMILARITY.
 FT DISULFID 1674 1687 BY SIMILARITY.
 FT DISULFID 1770 1782 BY SIMILARITY.
 FT DISULFID 1777 1791 BY SIMILARITY.
 FT DISULFID 1793 1806 BY SIMILARITY.
 FT DISULFID 1812 1824 BY SIMILARITY.
 FT DISULFID 1818 1833 BY SIMILARITY.
 FT DISULFID 1835 1847 BY SIMILARITY.

Query Match 16.6%; Score 110.5; DB 1; length 2871;
 Best Local Similarity 26.3%; Pred. No. 0.018; Indels 47; Gaps 7;

Matches 35; Conservative 10; Mismatches 41; Indels 47; Gaps 7;

QY 23 CT-----ECYCH-----CTGGADCTCTDCTACTG--CGCNPNAHTC-----55
 DB 1212 CTNSEGSYECSCQGFALMPDQRCSTIDECEDNPNICDGGCCTNIPGEYRCICDYGFMA 1271
 QY 56 -TDSKNVKAATCTGCTKCTNARTCTNSKCF-----EAKT-CTD-----ST 95
 DB 1272 SEDKKTCDVNECDLNPICLGTGFCENTKGTGTCGDMYSGRKKTGCTDINECEIGAH 1331
 QY 96 NCYKATCTNSTG 108
 DB 1332 NCDRHAVCTNAG 1344

RESULT 12
 MT_DREPO

ID MT_DREPO STANDARD; PRT; 73 AA.

AC 094550; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 01-NOV-1997 (Rel. 35, Last annotation update)

DE Metallothionein.

OS Dreissena polymorpha (Zebra mussel).

OC Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconcha; Veneroida;

OC Dreissenoida; Dreissenidae; Dreissena.

NCBI_Taxid=45954;

11

SEQUENCE FROM N.A.

RA Szecken S.R., Engelken J., Hildebrandt A.;

RL Submitted (Sep-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

SEQUENCE OF 2-40.

RA Szecken S.R., Engelken J., Hildebrandt A.;

RL Submitted (Jul-1997) to the SWISS-PROT data bank.

CC -1- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR

CC SENSITIZATION OF TOXIC METAL IONS (BY SIMILARITY).

CC -1- INDUCTION: BY CADMIUM.

CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U67347; AAB07548.1; -.

DR HSSP; P02795; ZMHU.

DR InterPro; IPRO01008; Metllthion_mollusc.

DR PRINTS; PR00875; MTMOLLUSC.

KW Metal-binding; Metal-thiolate cluster; Chelation; Cadmium.

SO SEQUENCE 73 AA; 7364 MW; DD3398705C4DAEE9 CRC64;

Query Match 16.4%; Score 109; DB 1; length 73;
 Best Local Similarity 28.6%; Pred. No. 0.0017;

Matches 20; Conservative 6; Mismatches 40; Indels 4; Gaps 1;

QY 28 CHCTGGADCTCTDCTGCGCNPNAHTCTDSEKNVKAATCTGCTKCTNARTCTNSKCFE 87

DB 5 CMCVETGDCRCADGSCSDSCSNC-----KCGDSCCKSKPCCGKNVYCKGCMCGCGVGTG 60

QY 88 AKTCTDSTNC 97

DB 61 PDSCTCDSC 70

RESULT 13

ITB7_HUMAN

ID ITB7_HUMAN STANDARD; PRT; 798 AA.

AC P26010; 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DE Integrin beta-7 precursor.

GN ITGB7.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_Taxid=9606;

11

SEQUENCE FROM N.A. (SHORT AND LONG FORMS).

RC TISSUE=Leukocyte;

RA MEDLINE=91250405; PubMed=2040616;

RT Erle D.J., Rueegg C., Sheppard D., Pytela R.;

RT "Complete amino acid sequence of an integrin beta subunit (beta 7)

RT identified in leukocytes.";

RL J. Biol. Chem. 266:11009-11016(1991).

12

SEQUENCE FROM N.A.

RA MEDLINE=91190778; PubMed=2083230;

RT Yuan Q., Jiang W.-M., Krissansen G.W., Watson J.D.;

RT "Cloning and sequence analysis of a novel beta 2-related integrin

RT transcript from T lymphocytes: homology of integrin cysteine-rich

RT repeats to domain III of laminin B chains.";

RL Int. Immunol. 2:1097-1108(1990).

13

REVISIONS.

RA MEDLINE=92135083; PubMed=1777426;

RT Yuan Q., Jiang W.-M., Krissansen G.W., Watson J.D.;

RT "Cloning and sequence analysis of a novel beta 2-related integrin

RT transcript from T lymphocytes: homology of integrin cysteine-rich

RT repeats to domain III of laminin B chains.";

RL Int. Immunol. 3:1373-1374(1991).

14

SEQUENCE FROM N.A.

RA MEDLINE=93002753; PubMed=1382574;

RT Jiang W.-M., Jenkins D., Yuan Q., Leung E., Choo K.H., Watson J.D.,

RA Krissansen G.W.;

RT "The gene organization of the human beta 7 subunit, the common beta

RT subunit of the leukocyte integrins HML-1 and IPAM-1.";

RL Int. Immunol. 4:1031-1040(1992).

15

MUTAGENESIS OF ASP-159.

RA MEDLINE=20400502; PubMed=10837471;

RT Higgins J.M.G., Cernadas M., Tan K., Irlie A., Wang J.-H., Takada Y.,

RA Brenner M.B.;

RT "The role of alpha and beta chains in ligand recognition by beta 7

RT integrins.";

RL J. Biol. Chem. 275:25652-25664(2000).

CC -1- FUNCTION: INTEGRIN ALPHA-4/BETA-7 (PEYER'S PATCHES-SPECIFIC HOMING

CC RECEPTOR LPAM-1) IS EXPECTED TO PLAY A ROLE IN ADHESIVE

CC INTERACTIONS OF LEUKOCYTES. IT IS A RECEPTOR FOR FIBRONECTIN AND

CC RECOGNIZES ONE OR MORE DOMAINS WITHIN THE ALTERNATIVELY SPLICED

CC CS-1 REGION OF FIBRONECTIN. INTEGRIN ALPHA-4/BETA-7 IS ALSO A

CC RECEPTOR FOR MADCAM1 AND VCAM1. IT RECOGNIZES THE SEQUENCE L-D-T

CC IN MADCAM1. INTEGRIN ALPHA-E/BETA-7 (HML-1) IS A RECEPTOR FOR E-

CC CADHERIN.

CC -1- SUBUNIT: HETERO-DIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-7

FT	REPEAT	566	604	III.
FT	REPEAT	605	640	IV.
FT	DISULFID	45	476	BY SIMILARITY.
FT	DISULFID	51	61	BY SIMILARITY.
FT	DISULFID	54	91	BY SIMILARITY.
FT	DISULFID	64	80	BY SIMILARITY.
FT	DISULFID	216	223	BY SIMILARITY.
FT	DISULFID	271	311	BY SIMILARITY.
FT	DISULFID	412	428	BY SIMILARITY.
FT	DISULFID	448	688	BY SIMILARITY.
FT	DISULFID	474	478	BY SIMILARITY.
FT	DISULFID	488	500	BY SIMILARITY.
FT	DISULFID	497	537	BY SIMILARITY.
FT	DISULFID	502	511	BY SIMILARITY.
FT	DISULFID	513	527	BY SIMILARITY.
FT	DISULFID	543	548	BY SIMILARITY.
FT	DISULFID	545	574	BY SIMILARITY.
FT	DISULFID	550	559	BY SIMILARITY.
FT	DISULFID	561	566	BY SIMILARITY.
FT	DISULFID	580	585	BY SIMILARITY.
FT	DISULFID	582	613	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	598	605	BY SIMILARITY.
FT	DISULFID	619	624	BY SIMILARITY.
FT	DISULFID	621	666	BY SIMILARITY.
FT	DISULFID	626	635	BY SIMILARITY.
FT	DISULFID	638	641	BY SIMILARITY.
FT	DISULFID	645	654	BY SIMILARITY.
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	279	279	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	434	434	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	477	477	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	531	531	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	590	590	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	674	674	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MOD_RES	778	778	PHOSPHORYLATION (BY TYR-KINASES) (BY SIMILARITY).
FT	VASPELIC	501	648	MISSING (IN SHORT ISOFORM).
FT	MURAGEN	159	159	D-3: LOSS OF INTEGRIN ALPHA-E/BETA-7 BINDING TO E-CADHERIN AND OF INTEGRIN ALPHA-4/BETA-7 BINDING TO MADCAM1.
FT	SEQUENCE	798 AA;	86903 MW;	CBB275DE0992385 CRC64;
QY	Query Match	16.4%;	Score 109;	DB 1; Length 798;
DB	Best Local Similarity	32.0%;	Pred. 0.0098;	
	Matches 32;	Conservative 5;	Mismatches 47;	Indels 16; Gaps 5;
QY	21 CIGCTGCGH-----CTGGADCTGCTD-----ACGCGCNC-P-NMHTCTDSKNCYKAATCT	68		
DB	560 CQCGVCHGHANHTGRACCSGDMDCISPEGGLCSGHRCKNCNRCCIDGYGALCDQCP	639		
QY	69 G-STKCNFARTCTNSKDCFEAKTCTDSTNCKYKATCTACTST	107		
DB	640 GCKTPCERHRDCA---ECGAFRTGLALNCSYACAHNTVT	676		
RESULT 14				
C170_GIALA				
ID	C170_GIALA	STANDARD;	PRT;	328 AA.
AC	PI5799;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Surface antigen CRP170 (Fragment).			
OS	Giardia lamblia (Giardia intestinalis).			
OC	Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.			
OX	NCBI_TaxID=5741;			
BN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 30957 / WB.			
RA	MEDLINE=88089405; Pubmed=3335828;			
RA	Adem R.D., Aggarwal A., Lal A.A., de la Cruz V.F., McCutchan T.			

RA Nash T.E.;
 RT "Antigenic variation of a cysteine-rich protein in Giardia lamblia."; J. Exp. Med. 167:109-118(1988).
 CC -1- MISCELLANEOUS: CYSTEINE-RICH, ANTIGENICALLY VARIANT SURFACE PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X06741; CAA29916.1; -
 DR PIR: S00530; S00530.
 DR InterPro: IPR002174; Furin-like.
 DR SMART: SM00261; FU; 3.
 DR Repeat: Antigen.
 FT NON_TER 1 1
 FT REPEAT 38 102
 FT REPEAT 103 167
 FT NON_TER 328 328
 FT SEQUENCE 328 AA; 33438 MW; 373A697A30EDCA21 CRC64;
 SO
 Query Match 16.3%; Score 108.5; DB 1; Length 328;
 Best Local Similarity 28.5%; Pred. No. 0.0057;
 Matches 37; Conservative 9; Mismatches 41; Indels 43; Gaps 7;
 Oy 23 CTECYCHCTGGADCTGCTDA-----CTGC-----GNCPPNA---HT---CT 56
 Db 167 CAEGCKPMTAGTGCFCSDANCERCDDNDVCARCSGAPENGKCPATPCGCHSCGCGT 226
 Oy 57 DSKNCVKAATCTG-----STGCNTRFCTCNKDCFEAKCTCDSTN-CYKA 100
 Db 227 ENAMTNQADKCTGCKEGRYLKPESAAQSGALTAEBCTSDKTHFTRKAGDSKMGCLSC 286
 Oy 101 TACTNS-TGC 109
 Db 287 SDATHGTTGC 296
 Db
 RESULT 15
 FBNI_MOUSE STANDARD; PRT; 2871 AA.
 ID FBNI_MOUSE
 AC 061554; 060826;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Fibrillin 1 precursor.
 GN FBNI OR FBNI-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95130561; PubMed=7829516;
 RA Yin W., Germiller J., Sanguinetti C., Smiley E., Pangillan T., Pereira L., Ramirez F., Bonadio J.;
 RA "Primary structure and developmental expression of Fbn-1, the mouse fibrillin gene."; J. Biol. Chem. 270:1798-1806(1995).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Kidney;
 RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
 RA Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE

CC MICROFIBRILS (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L29454; AAA56840.1; -
 DR EMBL: U22493; AAA64217.1; -
 DR HSSP: P35555; IAP1.
 DR MGD: MGI:95489; FBN1.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002212; TB.
 DR Pfam: PF00008; EGF; 46.
 DR Pfam: PF00683; TB; 9.
 DR SMART: SM00179; EGF_CA; 42.
 DR SMART: SM00001; EGF-like; 4.
 DR PROSITE: PS00010; ASX_HYDROXYL; 43.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 38.
 DR PROSITE: PS01187; EGF_CA; 45.
 DR Extracellular matrix; Calcium-binding; glycoprotein; EGF-like domain; Kunitz domain; Signal; Multigene family.
 KW Repeat: Signal; Multigene family.
 FT CHAIN 1 27
 FT SIGNAL 1 27
 FT DOMAIN 81 2871
 FT DOMAIN 115 112
 FT DOMAIN 147 178
 FT DOMAIN 246 287
 FT DOMAIN 288 329
 FT DOMAIN 330 401
 FT DOMAIN 402 446
 FT DOMAIN 449 489
 FT DOMAIN 490 529
 FT DOMAIN 530 571
 FT DOMAIN 572 612
 FT DOMAIN 613 653
 FT DOMAIN 656 721
 FT DOMAIN 723 764
 FT DOMAIN 765 806
 FT DOMAIN 807 846
 FT DOMAIN 910 951
 FT DOMAIN 952 1018
 FT DOMAIN 1028 1069
 FT DOMAIN 1070 1112
 FT DOMAIN 1113 1154
 FT DOMAIN 1155 1196
 FT DOMAIN 1197 1237
 FT DOMAIN 1238 1279
 FT DOMAIN 1280 1321
 FT DOMAIN 1322 1362
 FT DOMAIN 1363 1403
 FT DOMAIN 1404 1445
 FT DOMAIN 1446 1486
 FT DOMAIN 1487 1527
 FT DOMAIN 1528 1569
 FT DOMAIN 1569 1606
 FT DOMAIN 1606 1647
 FT DOMAIN 1648 1688
 FT DOMAIN 1689 1758
 FT DOMAIN 1759 1807
 FT DOMAIN 1808 1848
 FT DOMAIN 1849 1890
 FT DOMAIN 1891 1929
 FT DOMAIN 1930 1972
 FT DOMAIN 1973 2012
 FT FIBRILLIN 1, NON-CALCIUM BINDING.
 EGF-LIKE 1, NON-CALCIUM BINDING.
 EGF-LIKE 2, NON-CALCIUM BINDING.
 EGF-LIKE 3, NON-CALCIUM BINDING.
 EGF-LIKE 4, CALCIUM-BINDING.
 EGF-LIKE 5, CALCIUM-BINDING.
 TGFBRP 1.
 PRO-RICH.
 EGF-LIKE 6, NON-CALCIUM BINDING.
 EGF-LIKE 7, CALCIUM-BINDING.
 EGF-LIKE 8, CALCIUM-BINDING.
 EGF-LIKE 9, CALCIUM-BINDING.
 EGF-LIKE 10, CALCIUM-BINDING.
 TGFBRP 2.
 EGF-LIKE 11, CALCIUM-BINDING.
 EGF-LIKE 12, CALCIUM-BINDING.
 EGF-LIKE 13, CALCIUM-BINDING.
 EGF-LIKE 14, CALCIUM-BINDING.
 TGFBRP 3.
 EGF-LIKE 15, CALCIUM-BINDING.
 EGF-LIKE 16, CALCIUM-BINDING.
 EGF-LIKE 17, CALCIUM-BINDING.
 EGF-LIKE 18, CALCIUM-BINDING.
 EGF-LIKE 19, CALCIUM-BINDING.
 EGF-LIKE 20, CALCIUM-BINDING.
 EGF-LIKE 21, CALCIUM-BINDING.
 EGF-LIKE 22, CALCIUM-BINDING.
 EGF-LIKE 23, CALCIUM-BINDING.
 EGF-LIKE 24, CALCIUM-BINDING.
 EGF-LIKE 25, CALCIUM-BINDING.
 EGF-LIKE 26, CALCIUM-BINDING.
 TGFBRP 4.
 EGF-LIKE 27, CALCIUM-BINDING.
 EGF-LIKE 28, CALCIUM-BINDING.
 TGFBRP 5.
 EGF-LIKE 29, CALCIUM-BINDING.
 EGF-LIKE 30, CALCIUM-BINDING.
 EGF-LIKE 31, CALCIUM-BINDING.
 EGF-LIKE 32, CALCIUM-BINDING.
 EGF-LIKE 33, CALCIUM-BINDING.
 EGF-LIKE 34, CALCIUM-BINDING.

FT	DOMAIN	2013	2054	EGE-LIKE 35,	CALCIUM-BINDING
FT	DOMAIN	2055	2121	TEGBP 6,	
FT	DOMAIN	2127	2165	EGE-LIKE 36,	CALCIUM-BINDING
FT	DOMAIN	2206	2205	EGE-LIKE 37,	CALCIUM-BINDING
FT	DOMAIN	2206	2246	EGE-LIKE 38,	CALCIUM-BINDING
FT	DOMAIN	2247	2230	EGE-LIKE 39,	CALCIUM-BINDING
FT	DOMAIN	2291	2332	EGE-LIKE 40,	CALCIUM-BINDING
FT	DOMAIN	2333	2400	TEGBP 7,	
FT	DOMAIN	2402	2443	EGE-LIKE 41,	CALCIUM-BINDING
FT	DOMAIN	2444	2484	EGE-LIKE 42,	CALCIUM-BINDING
FT	DOMAIN	2485	2523	EGE-LIKE 43,	CALCIUM-BINDING
FT	DOMAIN	2524	2566	EGE-LIKE 44,	CALCIUM-BINDING
FT	DOMAIN	2567	2607	EGE-LIKE 45,	CALCIUM-BINDING
FT	DOMAIN	2607	2647	EGE-LIKE 46,	CALCIUM-BINDING
FT	DOMAIN	2648	2687	EGE-LIKE 47,	CALCIUM-BINDING
FT	DISULFID	85	94	BY SIMILARITY,	
FT	DISULFID	89	100	BY SIMILARITY,	
FT	DISULFID	102	111	BY SIMILARITY,	
FT	DISULFID	119	129	BY SIMILARITY,	
FT	DISULFID	123	134	BY SIMILARITY,	
FT	DISULFID	136	145	BY SIMILARITY,	
FT	DISULFID	150	160	BY SIMILARITY,	
FT	DISULFID	154	166	BY SIMILARITY,	
FT	DISULFID	168	177	BY SIMILARITY,	
FT	DISULFID	250	271	BY SIMILARITY,	
FT	DISULFID	257	271	BY SIMILARITY,	
FT	DISULFID	273	286	BY SIMILARITY,	
FT	DISULFID	292	304	BY SIMILARITY,	
FT	DISULFID	299	313	BY SIMILARITY,	
FT	DISULFID	315	328	BY SIMILARITY,	
FT	DISULFID	453	465	BY SIMILARITY,	
FT	DISULFID	460	474	BY SIMILARITY,	
FT	DISULFID	476	488	BY SIMILARITY,	
FT	DISULFID	494	504	BY SIMILARITY,	
FT	DISULFID	499	513	BY SIMILARITY,	
FT	DISULFID	515	528	BY SIMILARITY,	
FT	DISULFID	534	546	BY SIMILARITY,	
FT	DISULFID	541	555	BY SIMILARITY,	
FT	DISULFID	557	570	BY SIMILARITY,	
FT	DISULFID	576	587	BY SIMILARITY,	
FT	DISULFID	582	596	BY SIMILARITY,	
FT	DISULFID	598	611	BY SIMILARITY,	
FT	DISULFID	617	628	BY SIMILARITY,	
FT	DISULFID	623	637	BY SIMILARITY,	
FT	DISULFID	639	652	BY SIMILARITY,	
FT	DISULFID	727	739	BY SIMILARITY,	
FT	DISULFID	734	748	BY SIMILARITY,	
FT	DISULFID	750	763	BY SIMILARITY,	
FT	DISULFID	769	781	BY SIMILARITY,	
FT	DISULFID	776	790	BY SIMILARITY,	
FT	DISULFID	792	805	BY SIMILARITY,	
FT	DISULFID	811	821	BY SIMILARITY,	
FT	DISULFID	816	830	BY SIMILARITY,	
FT	DISULFID	832	845	BY SIMILARITY,	
FT	DISULFID	914	926	BY SIMILARITY,	
FT	DISULFID	921	935	BY SIMILARITY,	
FT	DISULFID	937	950	BY SIMILARITY,	
FT	DISULFID	1032	1044	BY SIMILARITY,	
FT	DISULFID	1039	1053	BY SIMILARITY,	
FT	DISULFID	1055	1068	BY SIMILARITY,	
FT	DISULFID	1074	1086	BY SIMILARITY,	
FT	DISULFID	1081	1095	BY SIMILARITY,	
FT	DISULFID	1097	1111	BY SIMILARITY,	
FT	DISULFID	1117	1129	BY SIMILARITY,	
FT	DISULFID	1124	1138	BY SIMILARITY,	
FT	DISULFID	1140	1153	BY SIMILARITY,	
FT	DISULFID	1159	1171	BY SIMILARITY,	
FT	DISULFID	1166	1180	BY SIMILARITY,	
FT	DISULFID	1182	1195	BY SIMILARITY,	
FT	DISULFID	1201	1212	BY SIMILARITY,	
FT	DISULFID	1208	1221	BY SIMILARITY,	
FT	DISULFID	1223	1236	BY SIMILARITY,	
FT	DISULFID	1242	1254	BY SIMILARITY,	

```

FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.

Query Match          16.3%; Score 108.5; DB 1; Length 2871;
Best Local Similarity 26.3%; Pred. No. 0.027; Mismatches 42; Indels 47; Gaps
Matches 35; Conservative 9;

QY 23 CT-----ECYCH-----CTGADCTSTCDACTG--CGNCPNAHTC----- 55
    ||      || |      ||      ||      ||      ||      ||      ||
Db 1212 CTNSDGSYECSQPGFALMPDRSCTDIQCEENPNICDGGOCTNI PGHYRCLCYDGFMA 12711

QY 56 -TDSKNCVKAATCGSGSHKCAFTARTCTNSMDCF-----EAKT-CTD-----ST 95
    || || || || : || || : || || |
Db 1272 SEDMKTCVDVNECDLNPNICLSGTCENTKGSEFICHODMGYSKKRGKTGCTDINECEIGAH 13311

QY 96 NCYKATACTNSTG 108
    || : ||| : |
Db 1332 NCGRHAVCTNMAG 1344

Search completed: October 21, 2002, 16:38:48
Job time : 16.1765 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:44 : Search time 36.2353 Seconds
(without alignments)
534.712 Million cell updates/sec

Title: US-10-032-658-11

Perfect score: 664

Sequence: 1 MAFKTCGFSKMLVIAIVM.....DSTNCYKATACINSTGCPGH 112

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	664	100.0	112	5	016120 tenebrio mo
2	659	99.2	112	5	09u746 tenebrio mo
3	612	92.2	112	5	09u747 tenebrio mo
4	603	90.8	124	5	016121 tenebrio mo
5	599	90.2	112	5	016119 tenebrio mo
6	599	90.2	124	5	09u744 tenebrio mo
7	592	89.2	112	5	09u748 tenebrio mo
8	487	87.8	112	5	09u745 tenebrio mo
9	487	73.3	148	5	016122 tenebrio mo
10	385	58.0	118	5	09nc97 dendroides
11	385	58.0	122	5	09nc94 dendroides
12	377.5	56.9	108	5	046351 dendroides
13	377.5	56.9	108	5	046352 dendroides
14	372	56.0	136	5	09ncr2 dendroides
15	364.5	54.9	123	5	09nc99 dendroides
16	361	54.4	109	5	046346 dendroides

17	359	54.1	121	5	09ncr1 dendroides
18	357.5	53.8	129	5	09ncr0 dendroides
19	357.5	53.8	148	5	09nc98 dendroides
20	329.5	49.6	98	5	09nc96 dendroides
21	327.5	49.3	91	5	09ncr3 dendroides
22	326	49.1	104	5	09ncr5 dendroides
23	325.5	49.0	96	5	096419 dendroides
24	138.5	20.9	313	5	024330 dictyostel
25	133	20.0	107	5	09ncg9 crassostrea
26	124.5	18.8	548	5	09g045 giardia lam
27	120.5	18.1	350	5	094389 leishmania
28	119.5	18.0	861	11	09qmw58 mus sp. mot
29	118.5	17.8	2189	5	09b105 elmeria ten
30	117.5	17.7	401	5	09nh87 giardia lam
31	117	17.6	862	11	099jc2 mus musculus
32	116.5	17.5	75	5	09u1n5 crassostrea
33	116	17.5	667	5	09sw11 giardia lam
34	113.5	17.1	548	5	021629 caenorhabdi
35	113.5	17.1	1174	11	099k58 mus musculu
36	112.5	16.9	100	5	096260 litorea 1
37	112.5	16.9	169	4	014564 homo sapien
38	112.5	16.9	1210	6	095nd4 felis silve
39	112	16.9	718	5	09b107 entamoeba h
40	112	16.9	1111	5	09xwd6 caenorhabdi
41	111.5	16.8	1698	5	094438 chironomus
42	111	16.7	966	5	022378 caenorhabdi
43	111	16.7	1202	11	p97607 rattus norv
44	110.5	16.6	969	4	096k66 homo sapien
45	110.5	16.6	2872	11	09wuh8 rattus norv

ALIGNMENTS

RESULT 1	ID	PRELIMINARY:	PRT:	112 AA.
016120	016120			
AC	016120:			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	THEMAL HYSTERESIS PROTEIN ISOFORM YL-2 (1-3) PRECURSOR.			
OS	Tenebrio molitor (Yellow mealworm).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;			
OC	Cucujiformia; Tenebrionidae; Tenebrio.			
OX	NCBI_Taxid=7067;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=FAIRBODY;			
RA	MEDLINE=94400482; PubMed=9285581;			
RT	Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;			
RT	"Hyperactive antifreeze protein from beetles.";			
RL	Nature 388:727-728(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=FAIRBODY;			
RA	MEDLINE=94400482; PubMed=10471292;			
RT	Liou Y.C., Thibault P., Walker V.K., Davies P.L.;			
RT	"A complex family of highly heterogeneous and internally repetitive			
RT	hyperactive antifreeze proteins from the beetle Tenebrio molitor.";			
RL	Biochemistry 38:11415-11424(1999).			
DR	EMBL; AF160495; AAB70751.1;			
DR	InterPro; IPR003460; AFP.			
KW	Pfam; PF02420; AFP; 8.			
FT	SIGNAL.			
FT	CHAIN	29	112	POTENTIAL.
FT	SEQUENCE	112 AA;	11667 MW;	5883E6DCDF0F805 CRC64;
FT	Best local Similarity	100.0%;	Pred. No. 4.3e-67;	

August 21, 1997

```

RL Biochemistry 38:11415-11424(1999).
DR EMBL, AF159115; AAD55257.1; -.
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 8.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM D-16.
SQ SEQUENCE 112 AA; 11591 MW; BA290B5DBB806C67 CRC64;

Query Match 92.2%; Score 612; DB 5; Length 112;
Best Local Similarity 92.0%; Pred. No. 2,8e-61;
Matches 103; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAFKTCGSKKMLVIAIVMCLCTECYCHGCGADCTSCPTACGCGCNCPNAHTCTDSK 60
Db 1 MAFKTCGSKKMLVIAIVMCLCTECYCHGCGADCTSCPTACGCGCNCPNAHTCTDSK 60
QY 61 CVKATCTGSTRKCNMTARTCTNSKDCFEAKTCTDSTNCKAKTACTNSTGCPGH 112
Db 61 CVKATCTGSTRKCNMTARTCTNSKDCFEAKTCTDSTNCKAKTACTNSTGCPGH 112

RESULT 4
016121 PRELIMINARY; PRT; 124 AA.
AC 016121;
DT 01-JAN-1998 (TREMblrel, 05, Created)
DT 01-JAN-1998 (TREMblrel, 19, Last annotation update)
DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-3 (3-8) PRCSUNSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RA MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RA MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL, AF160496; AAB70752.1; -.
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 9.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 124 THERMAL HYSTERESIS PROTEIN ISOFORM YL-3
SQ SEQUENCE 124 AA; 12993 MW; 0DEB87105613791A CRC64;

Query Match 90.8%; Score 603; DB 5; Length 124;
Best Local Similarity 83.9%; Pred. No. 3.2e-60;
Matches 104; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 MAFKTCGSKKMLVIAIVMCLCTECYCHGCGADCTSCPTACGCGCNCPNAHTCTDSK - 59
Db 1 MAFKTCGSKKMLVIAIVMCLCTECYCHGCGADCTSCPTACGCGCNCPNAHTCTDSK 60
QY 60 -----NCKAPDCTGSTRKCNMTARTCTNSKDCFEAKTCTDSTNCKAKTACTNSTG 108
Db 61 CVKATCTGSTRKCNMTARTCTNSKDCFEAKTCTDSTNCKAKTACTNSTG 120
QY 109 CPGH 112

```

Db 121 CPGH 124

RESULT 5
016119 PRELIMINARY; PRT; 112 AA.

AC 016119; PRT; 112 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-1 (2-14) PRECURSOR.

OS Tenebrio molitor (Yellow mealworm).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;

OC Cucujiformia; Tenebrionidae; Tenebrio.

OX NCBI_TaxID=7067;

RN 11

RP SEQUENCE FROM N.A.

RC TISSUE=FATBODY;

RX MEDLINE=97429943; PubMed=9285581;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;

RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF160494; AAB7050.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 8.

KW Signal.

FT SIGNAL 1 28

FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM YL-1

FT SEQUENCE 112 AA; 11558 MW; C06D7AB6401BC909 CRC64;

SO

Query Match 90.2%; Score 599; DB 5; Length 112;

Best Local Similarity 90.2%; Pred. No. 8.1e-60;

Matches 101; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

DB 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

QY 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTGCPGH 112

RESULT 6
090744 PRELIMINARY; PRT; 124 AA.

AC 090744; PRT; 124 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE THERMAL HYSTERESIS PROTEIN ISOFORM 5-15 PRECURSOR.

OS Tenebrio molitor (Yellow mealworm).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;

OC Cucujiformia; Tenebrionidae; Tenebrio.

OX NCBI_TaxID=7067;

RN 11

RP SEQUENCE FROM N.A.

RC TISSUE=FATBODY;

RX MEDLINE=99400482; PubMed=10471292;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;

RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF159114; AAD55256.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 8.

KW Signal.

FT SIGNAL 1 28

FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9.

FT SEQUENCE 112 AA; 11532 MW; 8A6124CD31D0E19 CRC64;

SO

Query Match 89.2%; Score 592; DB 5; Length 112;

Best Local Similarity 89.3%; Pred. No. 4.9e-59;

Matches 100; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

DB 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

QY 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTGCPGH 112

RESULT 8
090745 PRELIMINARY; PRT; 112 AA.

AC 090745; PRT; 112 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF159114; AAD55260.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 9.

KW Signal.

FT SIGNAL 1 28

FT CHAIN 29 124 THERMAL HYSTERESIS PROTEIN ISOFORM 5-15.

FT SEQUENCE 124 AA; 13005 MW; 8EC6871047F2890B CRC64;

SO

Query Match 90.2%; Score 599; DB 5; Length 124;

Best Local Similarity 83.1%; Pred. No. 8.9e-60;

Matches 103; Conservative 3; Mismatches 6; Indels 12; Gaps 1;

QY 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 53

DB 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

QY 54 -----TCTDSKNCVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTG 108

DB 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTG 120

QY 109 CPGH 112

DB 121 CPGH 124

RESULT 7
090748 PRELIMINARY; PRT; 112 AA.

AC 090748; PRT; 112 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE THERMAL HYSTERESIS PROTEIN ISOFORM 4-9 PRECURSOR.

OS Tenebrio molitor (Yellow mealworm).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;

OC Cucujiformia; Tenebrionidae; Tenebrio.

OX NCBI_TaxID=7067;

RN 11

RP SEQUENCE FROM N.A.

RC TISSUE=FATBODY;

RX MEDLINE=99400482; PubMed=10471292;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;

RT "A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor.";

RL Biochemistry 38:11415-11424(1999).

DR EMBL; AF159114; AAD55256.1; -.

DR InterPro; IPR003460; AFP.

DR Pfam; PF02420; AFP; 8.

KW Signal.

FT SIGNAL 1 28

FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9.

FT SEQUENCE 112 AA; 11532 MW; 8A6124CD31D0E19 CRC64;

SO

Query Match 89.2%; Score 592; DB 5; Length 112;

Best Local Similarity 89.3%; Pred. No. 4.9e-59;

Matches 100; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

DB 1 MAFKTCGFSKRWLVAVIWMCLCTECYCHCTGADCTCTDAGTCGCGCPNAHTCTDSKN 60

QY 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTGCPGH 112

DB 61 CVKATCTGSGTKCNTARTCTNSKDCFEAKTCTDSTNCTKATCTNSTGCPGH 112

RESULT 8
090745 PRELIMINARY; PRT; 112 AA.

AC 090745; PRT; 112 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM C-9 PRECURSOR.
OS Tenebrio molitor (yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159117; AAD55259.1; -
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
KW Signal.
FT SIGNAL. 1 28 POTENTIAL.
FT CHAIN 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM C-9.
SQ SEQUENCE 112 AA; 12002 MW; A4CA9DFB0AF81FB CRC64;
Query Match 87.8%; Score 583; DB 5; Length 112;
Best Local Similarity 87.5%; Pred. No. 5e-58;
Matches 98; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
QY 1 MAFKTCGFSKKMLVIAYIVMCLCTECYCHCTGADCTGCTGAGCNCNPNATCTDSKN 60
DB 1 MAFKTCGFSKKMLVIAYIVMCLCTECYCHCTGADCTGCTGAGCNCNPNATCTDSKN 60
QY 61 CVKATCTGCTKNTACTCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 112
DB 61 CVKATCTGCTKNTACTCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 112
Db 61 CVKATCTGCTKNTACTCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 112
RESULT 9
OL6122
ID 016122 PRELIMINARY; PRT; 148 AA.
AC 016122;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-4 (2-20) PRECURSOR.
OS Tenebrio molitor (yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 11.
KW Signal.
FT SIGNAL. 1 28 POTENTIAL.
FT CHAIN 29 148 THERMAL HYSTERESIS PROTEIN ISOFORM YL-4
SQ SEQUENCE 148 AA; 15311 MW; EF07B36BB8684365 CRC64;

Query Match 73.3%; Score 487; DB 5; Length 148;
Best Local Similarity 60.8%; Pred. No. 3.6e-47;
Matches 90; Conservative 5; Mismatches 17; Indels 36; Gaps 2;
QY 1 MAFKTCGFSKKMLVIAYIVMCLCTECYCHCTGADCTGCTGAGCNCNPNATCTDSKN 60
DB 1 MAFKTCGFSKKMLVIAYIVMCLCTECYCHCTGADCTGCTGAGCNCNPNATCTDSKN 60
QY 61 CVKATCTGCTKNTACTCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 112
DB 61 CVKATCTGCTKNTACTCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 112
QY 88 ---AKTCTDSTNCYKATCTACTNSTGCGH 112
DB 121 CTTATCTCTNSKNTACTCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 148
RESULT 10
Q9NC07
ID 09NC07 PRELIMINARY; PRT; 118 AA.
AC 09NC07;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ANTIFREEZE PROTEIN 12.
GN AFP-12.
OS Dendroidea canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroidea.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
proteins of the pyrochroid beetle Dendroidea canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179416; AAF6367.1; -
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 7.
DR Pfam; PF02420; AFP; 7.
SQ SEQUENCE 118 AA; 12162 MW; 7DCC9542E41BD11B CRC64;
Query Match 58.0%; Score 385; DB 5; Length 118;
Best Local Similarity 61.4%; Pred. No. 7.7e-36;
Matches 70; Conservative 9; Mismatches 21; Indels 14; Gaps 2;
QY 11 KVLVIAYIVMCLCTECYCHCTGADCTGCTGAGCNCNPNATCTDSKNYKAACTG 69
DB 4 KALIIISVLMCLCHEYAVAVCTGGPNCNACTTACTCINCIPNALLACTDSTNCLKAVTCTR 63
QY 70 STKCMATCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 110
DB 64 STKCMATCTNSKDCFEAKCTDSTNCYKATCTACTNSTGCGH 117
RESULT 11
Q9NCR4
ID 09NCR4 PRELIMINARY; PRT; 122 AA.
AC 09NCR4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ANTIFREEZE PROTEIN 5 (FRAGMENT).
GN AFP-5.
OS Dendroidea canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroidea.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;


```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidea; Dendroidea;
OX NCBI_TaxID=51100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroidea canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179414; AAF86365.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
FT NON_TER 1
SQ SEQUENCE 123 AA; 12882 MW; A5D92CEAD81B4DA5 CRC64;

Query Match 54.9%; Score 364.5; DB 5; Length 123;
Best Local Similarity 61.9%; Pred. No. 1.6e-33;
Matches 70; Conservative 8; Mismatches 20; Indels 15; Gaps 4;

QY 13 LVIAVIYK-CICTECYC-HCTGGADCTCTDAGCGCNCPCNAHT-CTDSKNCKYKAATCTG 69
   ||:::||||| | |||||:|||| | ||||| | ||||| | |||||
DB 2 LIIGFVLMSCLCHEVYAOECTGGSDCSSCYVACTNCQNCNALPACTDSTNCKRAFTCTP 61

QY 70 STKCNARTCTNSKCFE-----AKTCTDSTNCKYKATCTNSTGCP 110
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 STMCNGAVTCTDSDCENAEETCTGNTCTGTAKTCTGNTCTGNTCTGNTCTGNTCTG 114

```

Search completed: October 21, 2002, 16:39:38
 Job time : 38.2353 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:03 ; Search time 9.35294 Seconds
(without alignments)
285.020 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTCTACTACGCGCPNA 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	96.2	124	20	AA09281
2	126	96.2	124	20	AA09282
3	126	96.2	148	20	AA09280
4	124	94.7	24	20	AA09276
5	124	94.7	112	20	AA09278
6	124	94.7	112	20	AA09279
7	97	74.0	108	18	AA007693
8	97	74.0	109	18	AA007694
9	68	51.9	1679	22	AA007343
10	59	45.0	233	21	AA074791
11	58	44.3	24	17	AA098205

12	57	43.5	63	22	ABB66560
13	57	43.5	119	22	AAG91740
14	56.5	43.1	524	22	AA007370
15	56	42.7	1081	20	AA024319
16	55.5	42.4	225	21	AA036084
17	54	41.2	122	21	AA042599
18	53	40.5	1284	22	ABB61690
19	52	39.7	16	20	AA092275
20	52	39.7	727	11	AA005533
21	52	39.7	3457	15	AA062504
22	52	39.7	3457	15	AA062504
23	51.5	39.3	86	22	AB026420
24	51.5	39.3	1798	19	AA050896
25	51	38.9	59	21	AA057812
26	50.5	38.5	61	19	AA061601
27	50.5	38.5	61	21	AA057822
28	50.5	38.5	68	12	AA044774
29	50.5	38.5	68	13	AA025720
30	50.5	38.5	68	15	AA053383
31	50.5	38.5	68	22	AA078444
32	50.5	38.5	82	22	AA079428
33	50.5	38.5	215	21	AA056851
34	50.5	38.5	225	21	AA056850
35	50.5	38.5	230	21	AA056849
36	50.5	38.5	420	20	AA013362
37	50.5	38.5	420	20	AA005281
38	50.5	38.5	420	21	AA024396
39	50.5	38.5	420	21	AA085659
40	50.5	38.5	420	22	AA039899
41	50.5	38.5	420	22	AA012316
42	50.5	38.5	420	22	AA068594
43	50.5	38.5	420	22	AA080230
44	50.5	38.5	420	22	AA048106
45	50.5	38.5	420	22	AA048110

ALIGNMENTS

RESULT 1
ID AA09281 standard; Protein; 124 AA.
XX
AC AA09281;
XX
DT 06-JUL-1999 (first entry)
XX
DE YL-3 thermal hysteresis protein.
XX
KW Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
OS Tenebrio molitor.
XX
PN WO990493-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-CA000618.
XX
PR 26-JUN-1997; 97US-0882907.
XX
PA (TOOH) UNTV QUEENS KINGSTON.
XX
PI Davies PL, Graham LA, Liou Y, Walker VK;
XX
DR WPI; 1999-095739/08.
XX
N-PSDB; AAX33468.
XX
PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage
XX

PS Disclosure; Page 66; 88pp; English.

XX The present invention describes a nucleic acid (I) encoding an antifreeze protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD; (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c) (i) binds specifically to antibodies raised against the Tenebrio molitor YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4 or (iii) has a sequence that includes at least one repeating unit of the consensus sequence CTXSSXCXAXT (I). (II), optionally present in an organism, are used to improve preservation of frozen foods (specifically to reduce formation of ice crystals following exposure to repeated freeze-thaw cycles) or viability of biological materials (e.g. organs, cells or extracts) stored or transported at low temperature, or more generally to depress the freezing point of any aqueous solution, particularly where this is applied to an organism. (I) is used to produce transgenic animals and plants (e.g. grapes, oilseed crops, grains, citrus fruit or sugar cane) having better tolerance to freezing. Fragments of (I) are used as probes to detect or isolate coding sequences, in diagnosis (for determining gene expression) and as inhibitors (e.g. to render pests more sensitive to cold). Ab are used to detect (II) is usual immunoassays. (II) from T. molitor have specific activity 100 times greater than that of fish antifreeze proteins. The present sequence represents a thermal hysteresis protein (THP) from the present invention.

CC Sequence 124 AA;

XX

Query Match 96.2%; Score 126; DB 20; Length 124;
Best Local Similarity 87.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGXA DCTGCTACTGCGCXPNA 24
||| ||||| ||||| ||||| |||||
Db 30 CTGGA DCTGCTACTGCGCXPNA 52

RESULT 2
AAV09282
ID AAV09282 standard; Protein; 124 AA.
XX
AC AAV09282;
XX
DT 06-JUL-1999 (first entry)
XX
DE 5-15 thermal hysteresis protein.
XX
KM Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
OS Tenebrio molitor.
XX
PN WO9900493-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-CA00618.
XX
PR 26-JUN-1997; 97US-0882907.
XX
PA (TOOH) UNIV QUEBENS KINGSTON.
XX
PI Davies PL, Graham LA, Liou Y, Walker VK;
XX
DR WPI: 1999-095739/08.
DR N-PSDB; AAX33469.
XX
PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
XX used to improve low temperature tolerance of fish, plants etc., and
XX quality of foods or biological materials during frozen storage
XX
PS Disclosure; Page 67-68; 88pp; English.
XX
CC The present invention describes a nucleic acid (I) encoding an antifreeze

CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD; (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c) (i) binds specifically to antibodies raised against the Tenebrio molitor YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4 or (iii) has a sequence that includes at least one repeating unit of the consensus sequence CTXSSXCXAXT (I). (II), optionally present in an organism, are used to improve preservation of frozen foods (specifically to reduce formation of ice crystals following exposure to repeated freeze-thaw cycles) or viability of biological materials (e.g. organs, cells or extracts) stored or transported at low temperature, or more generally to depress the freezing point of any aqueous solution, particularly where this is applied to an organism. (I) is used to produce transgenic animals and plants (e.g. grapes, oilseed crops, grains, citrus fruit or sugar cane) having better tolerance to freezing. Fragments of (I) are used as probes to detect or isolate coding sequences, in diagnosis (for determining gene expression) and as inhibitors (e.g. to render pests more sensitive to cold). Ab are used to detect (II) is usual immunoassays. (II) from T. molitor have specific activity 100 times greater than that of fish antifreeze proteins. The present sequence represents a thermal hysteresis protein (THP) from the present invention.

CC Sequence 124 AA;

XX

Query Match 96.2%; Score 126; DB 20; Length 124;
Best Local Similarity 87.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGXA DCTGCTACTGCGCXPNA 24
||| ||||| ||||| ||||| |||||
Db 30 CTGGA DCTGCTACTGCGCXPNA 52

RESULT 3
AAV09280
ID AAV09280 standard; Protein; 148 AA.
XX
AC AAV09280;
XX
DT 06-JUL-1999 (first entry)
XX
DE YL-4 thermal hysteresis protein.
XX
KM Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
OS Tenebrio molitor.
XX
PN WO9900493-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-CA00618.
XX
PR 26-JUN-1997; 97US-0882907.
XX
PA (TOOH) UNIV QUEBENS KINGSTON.
XX
PI Davies PL, Graham LA, Liou Y, Walker VK;
XX
DR WPI: 1999-095739/08.
DR N-PSDB; AAX33467.
XX
PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
XX used to improve low temperature tolerance of fish, plants etc., and
XX quality of foods or biological materials during frozen storage
XX
PS Disclosure; Page 64; 88pp; English.
XX
CC The present invention describes a nucleic acid (I) encoding an antifreeze protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD; (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c) (i) binds specifically to antibodies raised against the Tenebrio molitor

CC YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4
CC or (iii) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSSXXCXAXT (1). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific
CC activity 100 times greater than that of fish antifreeze proteins. The
CC present sequence represents a thermal hysteresis protein (THP) from the
CC present invention.
XX
SQ Sequence 148 AA;
Query Match 96.2%; Score 126; DB 20; Length 148;
Best Local Similarity 87.0%; Pred. No. 1.7e-07;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| ||||| ||||| |||||
Db 30 CTGAADCTCTXACTGCGXCPNA 52
RESULT 4
AA09276
ID AAY09276 standard; peptide: 24 AA.
XX
AC AAY09276;
XX
DT 06-JUL-1999 (first entry)
XX
DE Antifreeze protein amino acid sequence SEQ ID NO:4.
XX
KW Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
OS Synthetic.
OS Tenebrio molitor.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..24 /note= "X is unspecified"
XX
PN WO9900493-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-CA00618.
XX
PR 26-JUN-1997; 97US-0882907.
XX
PA (T00H) UNIV QUEBENS KINGSTON.
XX
PI Davies PL, Graham LA, Liou Y, Walker VK;
XX
DR WPI: 1999-095739/08.
XX
PS New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage
XX
XX Claim 14; Page 56; 88pp; English.
CC The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)

CC (i) binds specifically to antibodies raised against the Tenebrio molitor
CC YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4
CC or (iii) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSSXXCXAXT (1). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific
CC activity 100 times greater than that of fish antifreeze proteins.
XX
SQ Sequence 24 AA;
Query Match 94.7%; Score 124; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CTGXADCTCTXACTGCGXCPNA 24
||| ||||| ||||| |||||
Db 2 CTGXADCTCTXACTGCGXCPNA 24
RESULT 5
AA09278
ID AAY09278 standard; Protein; 112 AA.
XX
AC AAY09278;
XX
DT 06-JUL-1999 (first entry)
XX
DE YL-1 thermal hysteresis protein.
XX
KW Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
KW Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
OS Tenebrio molitor.
XX
PN WO9900493-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-CA00618.
XX
PR 26-JUN-1997; 97US-0882907.
XX
PA (T00H) UNIV QUEBENS KINGSTON.
XX
PI Davies PL, Graham LA, Liou Y, Walker VK;
XX
DR WPI: 1999-095739/08.
XX
PS N-PSDB: AAX3465.
XX
PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage
XX
XX Disclosure; Page 60; 88pp; English.
CC The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/mL; (c)
CC (i) binds specifically to antibodies raised against the Tenebrio molitor
CC YL1-YL4 antifreeze proteins, (ii) is at least 60% identical with YL1-4
CC or (iii) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXSSXXCXAXT (1). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically

CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more
CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific
CC activity 100 times greater than that of fish antifreeze proteins. The
CC present sequence represents a thermal hysteresis protein (THP) from the
CC present invention.
XX
SQ Sequence 112 AA;
Query Match 94.7%; Score 124; DB 20; Length 112;
Best Local Similarity 87.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGXADCTSCYXACTGCGXCPNA 24
||| ||||| ||||| ||||| |||||
Db 30 CTGADCTCTGACTGCGXCPNA 52
RESULT 6
AAV09279
ID AAV09279 standard; Protein; 112 AA.
XX
AC AAV09279;
XX
DT 06-JUL-1999 (first entry)
XX
DE YL-2 thermal hysteresis protein.
XX
KM Antifreeze; thermal hysteresis protein; THP; mealworm beetle;
XX Tenebrio molitor; absorption-inhibition; ice crystal; frozen food.
XX
OS Tenebrio molitor.
XX
PN W09900493-A1.
XX
PD 07-JAN-1999.
XX
PF 25-JUN-1998; 98WO-CA00618.
XX
PR 26-JUN-1997; 97US-0882907.
XX
PA (TGOH) UNIV QUEENS KINGSTON.
XX
PI Davies PL, Graham LA, Liou Y, Walker VK;
XX
DR WPI, 1999-095739/08.
DR N-PSDB; AAX33466.
XX
XX
PT New nucleic acid encoding antifreeze proteins of Tenebrio molitor -
PT used to improve low temperature tolerance of fish, plants etc., and
PT quality of foods or biological materials during frozen storage
XX
PS Disclosure; Page 62; 88pp; English.
XX
XX The present invention describes a nucleic acid (I) encoding an antifreeze
CC protein (II) that: (a) has calculated molecular weight (m.w.) of 7-13 kD;
CC (b) has thermal hysteresis activity (THA) over 1.5 deg. C at 1 mg/ml; (c)
CC (1) binds specifically to antibodies raised against the Tenebrio molitor
CC YL-IV4 antifreeze proteins, (II) is at least 60% identical with YL-4
CC or (III) has a sequence that includes at least one repeating unit of the
CC consensus sequence CTXXXXCXAXR (1). (II), optionally present in an
CC organism, are used to improve preservation of frozen foods (specifically
CC to reduce formation of ice crystals following exposure to repeated
CC freeze-thaw cycles) or viability of biological materials (e.g. organs,
CC cells or extracts) stored or transported at low temperature, or more

CC generally to depress the freezing point of any aqueous solution,
CC particularly where this is applied to an organism. (I) is used to
CC produce transgenic animals and plants (e.g. grapes, oilseed crops,
CC grains, citrus fruit or sugar cane) having better tolerance to freezing.
CC Fragments of (I) are used as probes to detect or isolate coding
CC sequences, in diagnosis (for determining gene expression) and as
CC inhibitors (e.g. to render pests more sensitive to cold). Ab are used to
CC detect (II) is usual immunoassays. (II) from T. molitor have specific
CC activity 100 times greater than that of fish antifreeze proteins. The
CC present sequence represents a thermal hysteresis protein (THP) from the
CC present invention.
XX
SQ Sequence 112 AA;
Query Match 94.7%; Score 124; DB 20; Length 112;
Best Local Similarity 87.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGXADCTSCYXACTGCGXCPNA 24
||| ||||| ||||| ||||| |||||
Db 30 CTGADCTCTGACTGCGXCPNA 52
RESULT 7
AAW07693
ID AAW07693 standard; Protein; 108 AA.
XX
AC AAW07693;
XX
DT 09-SEP-1997 (first entry)
XX
DE Dendroides canadensis thermal-hysteresis protein.
XX
KM THP; antifreeze; transformation; plant.
XX
OS Dendroides canadensis.
XX
FH Key
FT 1..19 Location/Qualifiers
FT Peptide
FT /label= Signal
FT Protein
FT /product= Thermal-hysteresis-protein
FT Region
FT /label= Repeat_A
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region
FT /label= Repeat_B
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region
FT /label= Repeat_C
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region
FT /label= Repeat_D
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region
FT /label= Repeat_E
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region
FT /label= Repeat_F
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region
FT /label= Repeat_G
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence"
XX
PN W09640973-A1.
XX
PD 19-DEC-1996.
XX
PF 05-JUN-1996; 96WO-US08815.
XX
PR 07-JUN-1995; 95US-0485359.
PR 08-DEC-1995; 95US-0569594.

```

XX (UNOT ) UNIV NOTRE DAME DU LAC.
PA
XX
XX Duman JG;
PI
XX WPI: 1997-052352/05.
DR N-PSDB; AAT47150.
XX
PT DNA encoding Dendroides sp. thermal hysteresis protein - produces
PT protein having antifreeze properties, useful to protect plant cells,
PT seeds or plants from frost damage
XX
XX Disclosure: Page 31-32; 46pp; English.
XX
XX The present sequence represents a novel peptide which has antifreeze
XX properties. The nucleic acid sequence was derived from Dendroides
XX canadensis, and this protein is an example of a thermal-hysteresis
XX protein. The mature protein sequence contains a 13 amino acid sequence
XX and a 12 amino acid sequence that are repeated three and four times
XX respectively (see features table). The nucleic acid sequence can
XX be used to transform a plant cell for the heterologous production of
XX Dendroides thermal hysteresis protein (THP), i.e. to protect it from
XX frost damage. The DNA and polynucleotide sequences may be used as probes
XX for the isolation of THP coding sequences, from other organisms.
XX
SQ Sequence 108 AA;

Query Match 74.0%; Score 97; DB 18; Length 108;
Best Local Similarity 65.2%; Pred. No. 0.00026;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGXADCTCTYACTGCGXCPNA 24
|||::|||::|||::|||
DB 27 CTGSGDCRCTVSCCTDCQNCNPA 49

RESULT 8
AAW07694
ID AAW07694 standard; Protein: 109 AA.
XX
XX AAW07694;
AC
XX 09-SEP-1997 (first entry)
DT
XX
DE Dendroides canadensis thermal-hysteresis protein.
XX
XX THP; antifreeze; transformation; plant.
KW
XX Dendroides canadensis.
OS
XX
FH Key Location/Qualifiers
FT 1..19
FT Peptide /label= Signal
FT 20..108
FT Protein /product= Thermal-hysteresis_protein
FT Region 20..32
FT /label= Repeat_A
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 33..45
FT /label= Repeat_B
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 46..58
FT /label= Repeat_C
FT /note= "See AAW07696 for 13 amino acid consensus
FT sequence" Region 59..70
FT /label= Repeat_D
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 7182..82
FT /label= Repeat_E
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence" Region 83..94
FT /label= Repeat_F
FT /note= "See AAW07697 for 12 amino acid consensus

```

```

FT sequence" Region 95..104
FT /label= Repeat_G
FT /note= "See AAW07697 for 12 amino acid consensus
FT sequence"
XX
XX WO9640973-A1.
XX
XX 19-DEC-1996.
XX
XX 05-JUN-1996; 96WO-US08815.
XX
XX 07-JUN-1995; 95US-0485359.
XX PR 08-DEC-1995; 95US-0569594.
XX
XX (UNOT ) UNIV NOTRE DAME DU LAC.
PA
XX Duman JG;
PI
XX WPI: 1997-052352/05.
DR N-PSDB; AAT47151.
XX
XX DNA encoding Dendroides sp. thermal hysteresis protein - produces
PT protein having antifreeze properties, useful to protect plant cells,
PT seeds or plants from frost damage
XX
XX Disclosure: Page 33; 46pp; English.
XX
XX The present sequence represents a novel peptide which has antifreeze
XX properties. The nucleic acid sequence was derived from Dendroides
XX canadensis, and this protein is an example of a thermal-hysteresis
XX protein. The mature protein sequence contains a 13 amino acid sequence
XX and a 12 amino acid sequence that are repeated three and four times
XX respectively (see features table). The nucleic acid sequence can
XX be used to transform a plant cell for the heterologous production of
XX Dendroides thermal hysteresis protein (THP), i.e. to protect it from
XX frost damage. The DNA and polynucleotide sequences may be used as probes
XX for the isolation of THP coding sequences, from other organisms.
XX
SQ Sequence 109 AA;

Query Match 74.0%; Score 97; DB 18; Length 109;
Best Local Similarity 65.2%; Pred. No. 0.00027;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGXADCTCTYACTGCGXCPNA 24
|||::|||::|||::|||
DB 27 CTGSGDCRCTVSCCTDCQNCNPA 49

RESULT 9
AAU07343
ID AAU07343 standard; Protein: 1679 AA.
XX
XX AAU07343;
AC
XX 04-DEC-2001 (first entry)
DT
XX
DE 1-aminocyclopropane carboxylate (ACPC) synthase #12.
XX
XX 1-aminocyclopropane carboxylate synthase; ACPC synthase; brain trauma;
KW excitatory neurotransmission; neurodegeneration; stroke; nerve damage;
KW neurodegenerative disease; Alzheimer's disease; depression; epilepsy;
KW alcohol abuse; cognitive function; memory; learning impairment; human.
XX
XX Homo sapiens.
OS
XX
XX WO200168879-A2.
XX
XX 20-SEP-2001.
XX
XX 14-MAR-2001; 2001WO-EP02857.
XX PF 14-MAR-2000; 2000US-0189086.
XX PR

```

```

PR 05-APR-2000; 2000US-0194702.
PA (FARB ) BAYER AG.
PI
PI Ramakrishnan S;
PI
PI WPI; 2001-550286/61.
DR
XX
PT Isolated polynucleotide encoding a human
PT 1-aminocyclopropane-carboxylate (ACPC) synthase, useful for treating
PT brain trauma and neurodegenerative disease (e.g. Alzheimer's disease,
PT depression, epilepsy) -
XX
XX
PS Claim 1; Page 231-237; 242pp; English.
XX
CC The invention relates to reagents and methods for regulating excitatory
CC neurotransmission, and to prevent neurodegeneration. The method involves
CC the use of an expression vector or a reagent that modulates the activity
CC of a 1-aminocyclopropane-carboxylate (ACPC) synthase polypeptide. The
CC reagent is useful for modulating the activity of an ACPC synthase in a
CC disease such as stroke, a nerve damage or a neurodegenerative disease.
CC The ACPC synthase polypeptide, polynucleotides and modulators are also
CC useful for creating brain trauma and neurodegenerative disease (e.g.
CC Alzheimer's disease, depression, epilepsy). The ACPC synthase modulators
CC are also useful for treating alcohol abuse and improve cognitive function
CC and memory of patients with learning impairment. The present sequence
CC represents the amino acid sequence of human 1-aminocyclopropane-
CC carboxylate (ACPC) synthase #12, used in the method of the invention.
XX
SQ Sequence 1679 AA;
Query Match 51.9%; Score 68; DB 22; Length 1679;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
OY 2 CTGAXDCTCTACTGCGXC 21
DB 1112 CCGTGCCTCTCCCTCTGCGCGC 1131
RESULT 10
AAV74791
ID AAV74791 standard; Protein; 233 AA.
XX
AC AAV74791;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 263 protein sequence SEQ ID NO:1056.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US009346.
XX
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

```

```

XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M,
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
PI Tettelin H, Venter JC;
XX
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AA253553.
XX
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX
PS Claim 2; Page 606; 1453pp; English.
XX
CC AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 233 AA;
Query Match 45.0%; Score 59; DB 21; Length 233;
Best Local Similarity 45.8%; Pred. No. 11;
Matches 11; Conservative 0; Mismatches 9; Indels 4; Gaps 1;
OY 2 CTGAXDCTCTCT---XACTGCGXC 21
DB 159 CTGGCAACCTGTGTGCACTACTGTC 182
RESULT 11
AAR98205
ID AAR98205 standard; Protein; 24 AA.
XX
AC AAR98205;
XX
DT 30-DEC-1996 (first entry)
XX
DE Primer for amplifying MSH/MOMULV chimeric protein coding sequence.
XX
KW MoJoney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
KW polytropic MX27 provirus; targeted drug delivery; gene therapy;
KW single chain antibody; envelope protein; ss.
XX
OS Synthetic.
XX
PN WO9630504-A1.
XX
PD 03-OCT-1996.
XX
PF 22-MAR-1996; 96WO-US03908.
XX
PR 24-MAR-1995; 95US-0409648.
XX
PA (GENE-) GENETIC THERAPY INC.
PA (UTSC-) UNIV SOUTHERN CALIFORNIA.
XX
PI Anderson W, Chiang YL, Januszewski M, Mackrell AJ;
PI Zhao Y;
XX
XX
DR WPI; 1996-455352/45.
XX
PT Cell-targeted retroviral vector particles - having envelope protein
PT modified with targeting polypeptide

```

XX Example 2; Page 36; 73pp; English.
 PS
 CC Cell targeted retroviral vector particles can be used in gene
 CC therapy to deliver a heterologous gene to a target cell for
 CC expression of a heterologous polypeptide in that cell. The cell
 CC targeted retroviral vector particles comprise an envelope protein
 CC which is modified to contain a targeting polypeptide (a single chain
 CC antibody), or in the case of moloney murine leukaemia virus
 CC (MOMuLV), alpha melanotropin-stimulating hormone (MSH). Four primers
 CC (AAT33931, AAR98205 and AAT33932, AAR98206) were used to screen and
 CC sequence for synthetic MSH/MOMuLV chimeric protein coding sequences.
 CC Two other oligonucleotides used to substitute sequences in MOMuLV for
 CC MSH sequences are described in AAR98207 and AAR98208.
 CC
 XX Sequence 24 AA;
 SQ
 Query Match 44.3%; Score 58; DB 17; Length 24;
 Best Local Similarity 50.0%; Pred. No. 2.2;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
 QY 4 GXA^{CT}CT^{CT}ACTGCGXC 21
 Db 5 GCA^{CT}CT^{CT}GCGCACAC 22
 RESULT 12
 ABB66560
 ID ABB66560 standard; Protein; 63 AA.
 XX
 AC ABB66560;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 26472.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li FWD, Myers EW;
 PI
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL10663.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 26472; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 63 AA;
 DB
 Query Match 43.5%; Score 57; DB 22; Length 63;
 Best Local Similarity 45.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 2 CTGXA^{CT}CT^{CT}ACTGCGXC 21
 Db 20 CGGCGPCGGCGPCGGCGPC 39
 RESULT 13
 AAG91740
 ID AAG91740 standard; Protein; 119 AA.
 XX
 AC AAG91740;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 5494.
 XX
 KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOWA) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI: 2001-376931/40.
 DR N-PSDB; AAH66959.
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17; SEQ ID NO: 5494; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 119 AA;
 DB
 Query Match 43.5%; Score 57; DB 22; Length 119;
 Best Local Similarity 50.0%; Pred. No. 10;
 Matches 10; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

QY 9 TSCYXACTGCG-----XCP 22
|||: |||||
Db 93 TSCSLCTGCGCKTVAVTCTP 112

RESULT 14
AAU07370

ID AAU07370 standard; Protein: 524 AA.

XX AC AAU07370;

XX DT 18-DEC-2001 (first entry)

XX DE G protein-coupled receptor.

XX Human: mental disorder; thyroid disease; renal failure; anorexia;
KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
KW depression; Parkinson's disease; Alzheimer's disease; viral infection;
KW Huntington's disease; human immunodeficiency virus; type 2 diabetes;
KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
KW ngPCR.

XX OS Homo sapiens.

XX PN W0200162924-A2.

XX PD 30-AUG-2001.

XX PE 23-FEB-2001; 2001WO-US05989.

XX PR 24-FEB-2000; 2000US-0184602.

XX PR 24-FEB-2000; 2000US-0184604.

XX PR 24-FEB-2000; 2000US-0184606.

XX PR 24-FEB-2000; 2000US-0184689.

XX PR 24-FEB-2000; 2000US-0184690.

XX PR 24-FEB-2000; 2000US-0184710.

XX PR 24-FEB-2000; 2000US-0184712.

XX PR 24-FEB-2000; 2000US-0184715.

XX PR 24-FEB-2000; 2000US-0184716.

XX PR 24-FEB-2000; 2000US-0184725.

XX PR 24-FEB-2000; 2000US-0184822.

XX PA (PHMA) PHARMACIA & UPJOHN CO.

XX PI Vogel G, Wood LS, Parodi LA, Lind P;

XX DR WPI: 2001-570632/64.

XX PT Novel nucleic acid and encoded ngPCR-X, used to screen for compounds

XX PT for use in the treatment of mental disorders, such as Alzheimer's

XX PT disease, or Parkinson's disease -

XX PS Claim 31; Page 178-179; 263pp; English.

XX The invention relates to novel isolated human G protein-coupled

XX receptors (ngPCR-X). The ngPCR-X can be used for screening compounds

XX which can be used to treat mental disorders, thyroid disease, renal

XX failure, inflammatory conditions such as Crohn's disease, rheumatoid

XX arthritis, autoimmune disorders, schizophrenia, migraine, stroke,

XX dementia, depression, Parkinson's disease, Alzheimer's disease, and

XX Huntington's disease. They may also be used for treating viral infections

XX such as human immunodeficiency virus (HIV), type 2 diabetes, obesity,

XX anorexia, hypotension, hypertension, thrombosis, myocardial infarction,

XX atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726

XX represent the amino acid sequences of novel human G protein-coupled

XX receptors, ngPCR-2031 to ngPCR-2140 respectively, as described in the

XX invention.

XX Sequence 524 AA;

XX Query Match 43.1%; Score 56.5; DB 22; Length 524;

Best Local Similarity 44.8%; Pred. NO. 40;
Matches 13; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 2 CTGXAD-----CTGCTXA---CTGCGKC 21

Db 172 CTGCATATCAGACTATTTAAAGCAGCGCC 200

RESULT 15
AAU24319

ID AAU24319 standard; Protein: 1081 AA.

XX AC AAU24319;

XX DT 16-SEP-1999 (first entry)

XX DE Mouse dephosphorylase inhibiting p91-like protein #2.

XX KW Dephosphorylase inhibiting protein; p91; tyrosine phosphatase SHP-1;

XX KW SHP-2; inositol-5-phosphate SHP; phosphorylating tyrosine;

XX KW immunoreceptor; immunomodulatory agent.

XX OS Mus sp.

XX PN JP11169184-A.

XX PD 29-JUN-1999.

XX PF 12-DEC-1997; 97JP-0362285.

XX PR 12-DEC-1997; 97JP-0362285.

XX PA (UYOK-) UNIV OKAYAMA.

XX DR WPI: 1999-422622/36.

XX DR N-PSDB: AAX8976.

XX PT New peptide - useful for inhibiting dephosphorylase

XX PS Claim 2; Page 15-17; 30pp; Japanese.

XX CC The present invention describes new proteins for inhibiting

XX CC dephosphorylase. The proteins can be combined with tyrosine phosphatase

XX CC SHP-1, SHP-2 or inositol-5-phosphatase SHP by phosphorylating tyrosine.

XX CC The present invention also describes an immunoreceptor comprising one

XX CC the above proteins, and DNA coding the above proteins. The new proteins

XX CC can be used as an immunomodulatory agent. The present sequence

XX CC represents a protein from the present invention.

XX SQ Sequence 1081 AA;

QY 2 CTGXADCTCTYXACTGCGXCPNA 24

Db 744 CAG---CAACC-CTGCGGACTAA 762

Search completed: October 21, 2002, 16:38:22

Job time : 10.3529 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:37:29 ; Search time 3.70588 Seconds
(without alignments)
158.185 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTCTACTGCGCPNA 24

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	74.0	108	1	US-08-485-359-2
2	97	74.0	108	1	US-08-569-594-2
3	97	74.0	108	5	PCT-US96-08815-2
4	97	74.0	109	1	US-08-485-359-4
5	97	74.0	109	1	US-08-569-594-4
6	97	74.0	109	5	PCT-US96-08815-4
7	71	54.2	134	2	US-08-977-767-3
8	70.5	53.8	120	3	US-08-508-761B-22
9	68.5	52.3	45	4	US-08-900-230-14
10	66.5	50.8	801	1	US-07-906-349A-6
11	66.5	50.8	1400	4	US-08-630-915A-37
12	65.5	50.0	1417	4	US-08-900-230-3
13	63.5	48.5	57	1	US-07-609-716-56
14	61	46.6	47	3	US-08-482-085B-91
15	60	45.8	45	4	US-08-900-230-17
16	60	45.8	45	4	US-08-900-230-8
17	59.5	45.4	45	4	US-08-900-230-11
18	59	45.0	45	4	US-08-900-230-45
19	59	45.0	3788	4	US-09-336-447A-76
20	58.5	44.7	50	4	US-08-900-230-58
21	58.5	44.7	54	1	US-08-279-058B-24
22	58.5	44.7	102	3	US-08-974-022-53
23	58.5	44.7	102	4	US-08-795-445A-53
24	58.5	44.7	102	4	US-08-795-447A-53
25	58.5	44.7	102	4	US-08-974-186-53
26	58.5	44.7	102	4	US-08-795-446B-53
27	58.5	44.7	143	4	US-08-990-823-112

28	58	44.3	45	4	US-08-900-230-16	Sequence 16, App1
29	57.5	43.9	341	2	US-08-209-521-11	Sequence 11, App1
30	57	43.5	52	1	US-07-609-716-68	Sequence 68, App1
31	57	43.5	55	3	US-08-476-509B-46	Sequence 46, App1
32	55.5	42.4	46	4	US-08-900-230-40	Sequence 40, App1
33	54	41.2	45	4	US-08-900-230-43	Sequence 43, App1
34	54	41.2	45	4	US-08-900-230-44	Sequence 44, App1
35	54	41.2	48	5	PCT-US96-01720-8	Sequence 8, App1
36	53.5	40.8	33	4	US-09-047-288-4	Sequence 4, App1
37	53.5	40.8	45	4	US-08-900-230-52	Sequence 52, App1
38	53.5	40.8	47	1	US-08-451-947-95	Sequence 95, App1
39	53.5	40.8	47	1	US-08-451-947-96	Sequence 96, App1
40	53.5	40.8	47	2	US-08-424-826A-95	Sequence 95, App1
41	53.5	40.8	47	2	US-08-424-826A-96	Sequence 96, App1
42	53.5	40.8	47	3	US-08-928-694-95	Sequence 95, App1
43	53.5	40.8	47	3	US-08-928-694-96	Sequence 96, App1
44	53.5	40.8	47	5	PCT-US91-06950-95	Sequence 95, App1
45	53.5	40.8	47	5	PCT-US91-06950-96	Sequence 96, App1

ALIGNMENTS

RESULT 1
US-08-485-359-2
; Sequence 2, Application US/08485359
; Patent No. 5627051
; GENERAL INFORMATION:
; APPLICANT: Duman, John G.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; DENDROIDES CANADENSIS ANTIFREEZE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,359
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lammert, Steven R.
; REGISTRATION NUMBER: 27653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7258
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Dendroides canadensis
; US-08-485-359-2

Query Match 74.0%; Score 97; DB 1; Length 108;
Best local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 2 CTGXADCTCTACTGCGCPNA 24
||| :|| ||| :||| |||||

Db 27 CTGSDCRSCTVSCDCCNCPNA 49

RESULT 2

US-08-569-594-2

Sequence 2, Application US/08569594

Patent No. 5633451

GENERAL INFORMATION:

APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING

TITLE OF INVENTION: DENDROIDES CANADENSIS ANTIFREEZE PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Barnes & Thornburg

STREET: 11 South Meridian

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/569,594

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Lammert, Steven R.

REGISTRATION NUMBER: 27653

REFERENCE/DOCKET NUMBER: 835910-25377

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 231-7258

TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Dendroides canadensis

US-08-569-594-2

Query Match 74.0%; Score 97; DB 1; Length 108;

Best Local Similarity 65.2%; Pred. No. 2.3e-05;

Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXADCTCTXACTGCGXCPNA 24

||| :|| ||| :||| ||||

Db 27 CTGSDCRSCTVSCDCCNCPNA 49

RESULT 3

PCT-US96-08815-2

Sequence 2, Application PC/TUS9608815

GENERAL INFORMATION:

APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING

TITLE OF INVENTION: DENDROIDES ANTIFREEZE PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Barnes & Thornburg

STREET: 11 South Meridian

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/08815

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Lammert, Steven R.

REGISTRATION NUMBER: 27653

REFERENCE/DOCKET NUMBER: 835910-27026

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 231-7258

TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Dendroides canadensis

PCT-US96-08815-2

Query Match 74.0%; Score 97; DB 5; Length 108;

Best Local Similarity 65.2%; Pred. No. 2.3e-05;

Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXADCTCTXACTGCGXCPNA 24

||| :|| ||| :||| ||||

Db 27 CTGSDCRSCTVSCDCCNCPNA 49

RESULT 4

US-08-485-359-4

Sequence 4, Application US/08485359

Patent No. 5627051

GENERAL INFORMATION:

APPLICANT: Duman, John G.

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING

TITLE OF INVENTION: DENDROIDES CANADENSIS ANTIFREEZE PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Barnes & Thornburg

STREET: 11 South Meridian

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46204

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485,359

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lammert, Steven R.

REGISTRATION NUMBER: 27653

REFERENCE/DOCKET NUMBER: 835910-25377

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 231-7258

TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 109 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

;
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Dendroides canadensis
; US-08-485-359-4

Query Match 74.0%; Score 97; DB 1; Length 109;
Best Local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTXACTGCGXCPCNA 24
||| :|| ||| :||| ||||
Db 27 CTGGSDCRCTVCTDQNCPCNA 49

RESULT 5
US-08-569-594-4
; Sequence 4, Application US/08569594
; Patent No. 5633451
; GENERAL INFORMATION:

; APPLICANT: Duman, John G.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: DENDROIDES CANADENSIS ANTIFREEZE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,594
; FILING DATE:

; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Lammett, Steven R.
; REGISTRATION NUMBER: 27653
; REFERENCE/DOCKET NUMBER: 835910-25377
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7258
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Dendroides canadensis
; US-08-569-594-4

Query Match 74.0%; Score 97; DB 1; Length 109;
Best Local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTXACTGCGXCPCNA 24
||| :|| ||| :||| ||||
Db 27 CTGGSDCRCTVCTDQNCPCNA 49

RESULT 6
PCT-US96-08815-4
; Sequence 4, Application PC/TUS9608815
; GENERAL INFORMATION:
; APPLICANT: Duman, John G.

;
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: DENDROIDES ANTIFREEZE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08815
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lammett, Steven R.
; REGISTRATION NUMBER: 27653
; REFERENCE/DOCKET NUMBER: 835910-27026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7258
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Dendroides canadensis
; PCT-US96-08815-4

Query Match 74.0%; Score 97; DB 5; Length 109;
Best Local Similarity 65.2%; Pred. No. 2.3e-05;
Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 CTGAXDCTCTXACTGCGXCPCNA 24
||| :|| ||| :||| ||||
Db 27 CTGGSDCRCTVCTDQNCPCNA 49

RESULT 7

US-08-977-767-3
; Sequence 3, Application US/08977767
; Patent No. 5972684
; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Greenwald, Sara
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith

Sequence 6, Application US/07906349A
Patent No. 5434064
GENERAL INFORMATION:
APPLICANT: Schlessinger, Joseph
APPLICANT: Skolnik, Edward Y.
APPLICANT: Margolis, Benjamin L.
TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES AND
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906,349A
FILING DATE: 30-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/643,237
FILING DATE: 18-JAN-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-349A-6

Query Match 50.8%; Score 66.5; DB 1; Length 801;
Best Local Similarity 55.0%; Pred. No. 0.55;
Matches 11; Conservative 1; Mismatches 5; Indels 3; Gaps 1;

Oy 2 CTGXADCTCTXACTGCGXC 21
||| | :|| || || |
Db 239 CTG---CCTCTGACAGCAGC 255

RESULT 11
US-08-630-915A-37
Sequence 37, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-37

Query Match 50.8%; Score 66.5; DB 4; Length 1400;
Best Local Similarity 46.7%; Pred. No. 0.88;
Matches 14; Conservative 1; Mismatches 8; Indels 7; Gaps 2;

Oy 2 CTGXADCTCTXACTGCGXC 24
||| ||: | | |||| | | |
Db 882 CTGCTCTTCAGAGAGCTCCTGCTAGCTTA 911

RESULT 12
US-08-900-230-3
Sequence 3, Application US/08900230
Patent No. 6329197
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.
TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHETICAL: NO

ANTI-SENSE: NO
US-08-900-230-3

Query Match 50.0%; Score 65.5; DB 4; Length 1417;
Best Local Similarity 52.0%; Pred. No. 1.2;
Matches 13; Conservative 1; Mismatches 6; Indels 5; Gaps 2;

QY 2 CTGACDCTCTACCTGC-GXC 21
DB 501 CAGCAGCTTACGCTGCTGCTGC 525

RESULT 13

US-07-609-716-56
Sequence 56, Application US/07609716
Patent No. 5514581
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609, 716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ. ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-609-716-56

Query Match 48.5%; Score 63.5; DB 1; Length 57;
Best Local Similarity 55.0%; Pred. No. 0.13;
Matches 11; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

QY 2 CTGACDCTCTACCTGC-GXC 21
DB 28 CCGGGCTACCTACTG-GAC 46

RESULT 14
US-08-482-085B-91
Sequence 91, Application US/08482085B
Patent No. 6018030
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Richardson, Charles
APPLICANT: Chambers, James
APPLICANT: Causey, Stuart
APPLICANT: Pollock, Thomas J.

APPLICANT: Cappello, Joseph
APPLICANT: Crissman, John W.
TITLE OF INVENTION: No. 6018030el Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US

ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482, 085B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927, 258
FILING DATE: 04-NOV-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114, 618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053, 049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175, 155
FILING DATE: 29-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-398-3249
TELEFAX: 415-781-1989
INFORMATION FOR SEQ. ID NO: 91:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-482-085B-91

Query Match 46.6%; Score 61; DB 3; Length 47;
Best Local Similarity 54.5%; Pred. No. 0.22;
Matches 12; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 2 CT-GACDCTCTAC-TGCGXC 21
DB 10 CTCGATCTACATGATGCACC 31

RESULT 15
US-08-900-230-17
Sequence 17, Application US/08900230
Patent No. 6329197
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.
TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
TITLE OF INVENTION: USUS THEREOF
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of The Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.

ZIP: 11036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,230
FILING DATE: 23-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-C/JPM/ADM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-900-230-17

Query Match 45.8%; Score 60; DB 4; Length 45;
Best Local Similarity 47.4%; Pred. NO. 0.28;
Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Qy 3 TGXADCTSCYXACTGCGXC 21
Db 26 TGGGCGATCTCAACAGAGGC 44

Search completed: October 21, 2002, 16:40:36
Job time : 5.70588 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:36:38 : Search time 4.76471 Seconds
(without alignments)
484.006 Million cell updates/sec

Title: US-10-032-658-4
Perfect score: 131
Sequence: 1 XCTGXADCTSCRYACTGCGXCPNA 24

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	52.7	1176	2 A49848	nitrite reductase
2	65	49.6	1104	2 JH0181	nitrite reductase
3	60	45.8	325	2 H71271	hypothetical prote
4	59	45.0	1044	2 T43155	nitrite reductase
5	59	45.0	1044	2 T43160	nitrite reductase
6	58	44.3	719	2 T00266	hypothetical prote
7	57	43.5	63	2 S25772	testis-specific pr
8	56.5	43.1	2135	2 T14602	variant-specific s
9	55.5	42.4	79	2 T17014	metallothionein-11
10	55.5	42.4	141	2 PC1094	trophozoite surfac
11	54	41.2	807	2 E83726	assimilatory nitri
12	53.5	40.8	230	2 A38346	ultra-high-sulfur
13	53	40.5	25	1 SMMR	metallothionein -
14	53	40.5	65	2 A38739	metallothionein -
15	53	40.5	166	2 G98188	ferredoxin-like pr
16	53	40.5	166	2 AB3098	hypothetical prote
17	53	40.5	390	2 F69086	pyruvate formate-1
18	52	39.7	972	2 A30363	glycoprotein GP330
19	52	39.7	4660	2 T42737	gp330 protein prec
20	51.5	39.3	72	2 T07073	metallothionein ty
21	51.5	39.3	73	2 T07109	metallothionein-11
22	51.5	39.3	74	2 T16979	metallothionein-11
23	51.5	39.3	78	2 A23219	metallothionein-11
24	51.5	39.3	178	2 A23219	high-cysteine chor
25	51.5	39.3	279	2 A81225	ferredoxin, 4Fe-4S
26	51.5	39.3	279	2 D81997	probable ferredoxi
27	51.5	39.3	787	2 T05617	hypothetical prote
28	51.5	39.3	1797	2 A55677	laminin beta-2 cha
29	51.5	39.3	1798	2 A53869	laminin beta-2 cha

30	51.5	39.3	2215	2 T00348	LRI1 protein - mou
31	51	38.9	60	2 B27490	metallothionein B
32	51	38.9	166	2 D95346	NapF ferredoxin co
33	51	38.9	208	2 E90926	probable oxidoredu
34	51	38.9	208	2 A85775	probable oxidoredu
35	51	38.9	208	2 B64925	probable oxidoredu
36	51	38.9	223	2 B38346	ultra-high-sulfur
37	51	38.9	1806	2 T23298	hypothetical prote
38	50.5	38.5	60	2 S31723	metallothionein -
39	50.5	38.5	61	2 A37425	metallothionein 2
40	50.5	38.5	62	2 S54335	metallothionein-2c
41	50.5	38.5	62	2 S54336	metallothionein-2a
42	50.5	38.5	68	2 B46034	metallothionein 3'
43	50.5	38.5	68	2 S44392	metallothionein 3'
44	50.5	38.5	68	2 UC6521	metallothionein II
45	50.5	38.5	72	2 T07105	metallothionein-11

ALIGNMENTS

RESULT 1
A49848
nitrite reductase (NADH) (EC 1.6.6.4) - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 12-Sep-1997
C:Accession: A49848
R:Exley, G.E.; Colandene, J.D.; Garrett, R.H.
J. Bacteriol. 175, 2379-2392, 1993
A:Title: Molecular cloning, characterization, and nucleotide sequence of nit-6, the s
A:Reference number: A49848; MUID:93224461
A:Accession: A49848
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1176 <EXL>
A>Note: sequence inconsistent with nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:129213, NCBIPI:129214)
C:Keywords: oxidoreductase

Query Match 52.7%; Score 69; DB 2; Length 1176;
Best Local Similarity 60.0%; Pred. No. 1.1;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 2 CTGXADCTSCRYACTGCGXC 21
DB 515 CTSLGDLKSCRYACTGCGGC 534

RESULT 2
JH0181
nitrite reductase (NADH) (EC 1.6.6.4), long form - Emeritella nidulans
C:Species: Emeritella nidulans, Aspergillus nidulans
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 01-Dec-2000
C:Accession: JH0181; PS0298; PS0299
R:Johnstone, I.L.; McCabe, P.C.; Greaves, P.; Gurr, S.J.; Cole, G.E.; Brow, M.A.D.; U
Gene 90, 181-197, 1990
A:Title: Isolation and characterisation of the crna-nia-niad gene cluster for nitrat
A:Reference number: JH0181; MUID:9038264
A:Accession: JH0181
A:Molecule type: DNA
A:Residues: 1-1104 <JH>
A:Accession: PS0298
A:Molecule type: mRNA
A:Residues: 1-1104 <J02>
A:Accession: PS0299
A:Molecule type: mRNA
A:Residues: 1-770, 'E' <J03>
C:Comment: This protein is required for the assimilation of nitrate.
C:Genetics:
A:Gene: niaA
A:Introns: 59/2; 109/3; 289/3; 493/2; 561/1; 718/3; 771/2
C:Keywords: alternative splicing; NAD; oxidoreductase

Query Match	49.6%	Score 65	DB 2	Length 1104
Best Local Similarity	60.0%	Pred. No. 3		
Matches	12	Conservative	0	Mismatches 8; Indels 0; Gaps 0
OY	2	CTGAXDCSTCTACTGCGXC	21	
Db	519	CTTIADVKSCTKAGTGGCGC	538	

RESULT 3
H71271
hypochemical protein TP0856 - syphilis spirochete
C:Species: *Treponema pallidum* subsp. *pallidum* (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: H71271
R:Reiser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin, J.; Khailak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Ueterbach, T.; McDaniel, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
S:Science 281, 375-388, 1998
A:Title: Complete genome sequence of *Treponema pallidum*, the syphilis spirochete.
A:Reference number: A71250; MUID:9832770
A:Accession: H71271
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1525 <COL>
A:Cross-references: GB:AE001256; GB:AE000520; NID:g332317; PIDN:AAC65828.1; PID:g332317
A:Experimental source: strain Nichols
A:Genetics:
A:Gene: TP0856

Query Match	Similarity	45.8%	Score	60	DB	2	Length	325
Best Local	Similarity	47.6%	Pred. No.	4.9				
Matches	10; Conservative	2;	Mismatches	9;	Indels	0;	Gaps	0;
OY	3	TEHADCTCTXACTGCGXCPN	23					
Db	233	TGEGCKCTGCGRGTGTCGN	253					

RESULT 4
T43155
nitrite reductase (EC 1.7.99.3) - yeast (*Pichia angusta*)
C:Species: *Pichia angusta*
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 20-Apr-2000
C:Accession: T43155
R:Silverio, J.M.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z22318
A:Accession: T43155
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1044 <SIV>
A:Cross-references: EMBL:AJ223294; NID:e1339592; PID:e1339594; PIDN:CAN11230.1
A:Note: the source is designated as *Pichia angusta*
C:Genetics:
A:Gene: YNI1
A:Note: expression of YNI1 and YNR1 is coordinately regulated
;Keywords: oxidoreductase

Query Match	45.0%	Score 59	DB 2	Length 1044
Best Local Similarity	50.0%	Pred. No. 13		
Matches	10	Conservative	1	Mismatches 9; Indels 0; Gaps 0
OY	2	CTGAXDCTSTYACGTGCXC	21	
		1: 11 11 11 11		
db	479	CSLADVKKCTKAGTACGCGC	498	

RESULT 5
T43160
nitrite reductase (EC 1.7.99.3) - yeast (*Pichia angusta*)
C:Species: *Pichia angusta*
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 20-Apr-2000

C:Accession: T43160
R:Ritto, N.; Avila, J.; Perez, M.; Gonzalez, C.; Silverio, J.M.
J. Biochem. 317, 89-95, 1996
A>Title: The genes YNII and YNRI, encoding nitrite reductase and nitrate reductase re
A:Reference number: Z22319
A:Accession: T43160
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1044 <EBRI>
A:Cross-references: EMBL:Z68122; NID:e986903; PID:e21611; PIDD:CAA92206.1
A:Experimental source: strain NYC 495
A>Note: the source is designated as *Pichia angusta*
C:Genetics:
A:Gene: YNII
C:Keywords: oxidoreductase

Query Match 45.0%; Score 59; DB 2; Length 1044;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 CGGADCTGTCITACGTGGGC 21
 |: || | | | | | |
Db 479 CSSLDVKKCKTACGTACGCC 498

RESULT 6
r00266
hypothetical protein KIA0534 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00266
R:Naase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The compo
A:Reference number: Z14086; MUID:98290545
A:Accession: r00266
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-719 <NAC>
A:Cross-references: EMBL:AB011106; NID:g3043591; PIDN:BA25460.1; PID:g3043592
A:Experimental source: brain
C:Genetics:
A:Note: KIA0534

Query Match	Score	DB	Length
Best Similarity	50.0%	Pred No. 14;	
Best Local	50.0%	Pred No. 14;	
Matches	10;	Mismatches	10;
		Indels	0;
		Gaps	0;

RESULT 7
 S25772
 testis-specific protein Mst64Da - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Aug-1999
 C:Accession: S25772; A56565
 R:Kuhn, R., Kuhn, C., Boersch, D., Glaetzer, K.H., Schaefer, U., Schaefer, M.
 Mech. Dev. 35, 143-151, 1991
 A>Title: A cluster of four genes selectively expressed in the male germ line of *Drosophila*
 A:Reference number: A56565; MUID:92102953
 A:Accession: S25772
 A:Molecule type: DNA
 A:Residues: 1-63 <KUH>
 A:Cross-references: EMBL:X67703; NID:q11072; PIN:CAA47937.1; PID:q11073
 A:Note: sequence extracted from NCBI backbone (NCBI:74219, NCBI:74224)
 C:Genetics:
 A:Gene: Mst64Da
 A:Cross-references: FlyBase:FBgn0004172
 A:Map position: 3
 A:Introns: 13/3

C:Superfamily: fruit fly testis-specific protein
C:Keywords: spermatogenesis; tandem repeat

Query Match 43.5%; Score 57; DB 2; Length 63;
Best Local Similarity 45.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 CTGXADCTCTCTACCTGCGC 21
| | | | | | | | | |
Db 20 CGGCGCGGCGCGCCGCGGCGC 39

RESULT 8

T14602

variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: T14602

R:Voss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998

A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum

A:Reference number: Z18158

A:Accession: T14602

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2135 <VOS>

A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1

C:Genetics:

A:Gene: varph17

Query Match 43.1%; Score 56.5; DB 2; Length 2135;
Best Local Similarity 57.9%; Pred. No. 40;
Matches 11; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 2 CT-GXADCTCTCTACCTGCGC 19
| | | | | | | | | |
Db 1133 CTSGSDCTPCACACCTTGG 1151

RESULT 9

T17014

metallothionein-like protein AMT1 - apple tree
C:Species: Malus domestica (apple tree)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C:Accession: T17014

R:Reid, S.J.; Ross, G.S.

Physiol. Plantarum 100, 183-189, 1997

A:Title: Up-regulation of two cDNA clones encoding metallothionein-like proteins in apple

A:Reference number: Z18652

A:Accession: T17014

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-79 <REI>

A:Cross-references: EMBL:U61973; NID:g1655850; PID:g1655851

A:Experimental source: apple flesh cortical tissue

C:Genetics:

A:Gene: AMT1

C:Superfamily: metallothionein

C:Keywords: metal binding

Query Match 42.4%; Score 55.5; DB 2; Length 79;
Best Local Similarity 40.0%; Pred. No. 6.3;
Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;

QY 2 CTGXADC---TSCTXACTGCGXCPN 23
| | | | | | | | | |
Db 5 CGGKCGGVCSCGSCGCGMAPD 29

RESULT 10

PC1294

trophozoite surface antigen - Giardia lamblia (strain Adelaide-1) (fragment)
C:Species: Giardia lamblia

C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 09-Sep-1997
C:Accession: PC1294

R:EY, P.L.; Mayrhofer, G.

Gene 129, 257-262, 1993

A:Title: Two genes encoding homologous 70-kDa surface proteins are present within ind

A:Reference number: PC1294; MUID:93314970

A:Accession: PC1294

A:Molecule type: DNA

A:Residues: 1-141 <EYP>

A:Cross-references: GB:M97488; NID:g159129; PID:g159130

C:Genetics:

A:Gene: tsa

C:Keywords: surface antigen

Query Match 42.4%; Score 55.5; DB 2; Length 141;
Best Local Similarity 41.7%; Pred. No. 9.1;
Matches 10; Conservative 1; Mismatches 6; Indels 7; Gaps 1;

QY 2 CTGXADCTSC-----TXACTGC 18
| | | | | | | | | |
Db 102 CTSSDCTTCLDGYKKSACTKC 125

RESULT 11

E83726

assimilatory nitrite reductase (subunit) nasD [imported] - Bacillus halodurans (strai

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: E83726

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E83726

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-807 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04332.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: nasD

Query Match 41.2%; Score 54; DB 2; Length 807;
Best Local Similarity 45.0%; Pred. No. 41;
Matches 9; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 CTGXADCTCTCTACCTGCGXC 21
| | | | | | | | | |
Db 438 CASVADVGACTAGRSCGTC 457

RESULT 12

A8346

ultra-high-sulfur keratin 1 - mouse

C:Species: Mus musculus (house mouse)

C>Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 24-Sep-1999
C:Accession: A8346

R:Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogel, G.

J. Biol. Chem. 265, 21375-21380, 1990

A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skl

A:Reference number: A8346; MUID:91065960

A:Accession: A8346

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 <WOO>

A:Cross-references: GB:M37759; NID:g200961; PIDN:AAA0106.1; PID:g200962

C:Superfamily: ultra-high-sulfur keratin

Query Match 40.8%; Score 53.5; DB 2; Length 230;
Best Local Similarity 39.1%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 9; Indels 3; Gaps 1;

QY 2 CTGXADCTSC---CTXACTGCGXC 21

Db 55 CVPVCSGSCGGCGSSCGGCGSC 77

RESULT 13

SMR
metallothionein - cultivated mushroom
C:Species: *Agaricus bisporus* (cultivated mushroom)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 13-Sep-1996
C:Accession: A03286
R:Munger, K.; Lerch, K.
Biochemistry 24, 6751-6756, 1985
A:Title: Copper metallothionein from the fungus *Agaricus bisporus*: chemical and spectroscopic
A:Reference number: A03286
A:Accession: A03286
A:Molecule type: Protein
A:Residues: 1-25 <MUN>
A:Experimental source: mycelium; strain A-32
C:Comment: In contrast to the vertebrate metallothioneins, the fungal proteins bind copper
C:Comment: The absorptive, luminescent, and stereoptical properties of the copper MR are
C:Superfamily: metallothionein
C:Keywords: chelation; metal binding; metal-thiolate cluster

Query Match	40.5%;	Score 53;	DB 1;	Length 25;
Best Local Similarity	45.5%;	Pred. No. 5.7;		
Matches 10;	Conservative 3;	Mismatches 3;	Indels 6;	Gaps 2;

```
QY      2 CTGXADCT---SCTXACTGCG 19
        | : | : | | | | | |
Db      5 CSGASSCTCAGGCT--CSGCG 24
```

RESULT 14

A38739
metallothionein - sea urchin (*Strongylocentrotus purpuratus*)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 13-Sep-1991 #sequence_revision 13-Sep-1991 #text_change 20-Aug-1999
C:Accession: A38739
J:Nemer, M.; Thornton, R.D.; Stuebning, E.W.; Hatlow, P.
J. Biol. Chem. 266, 6586-6593, 1991
A:Title: Structure, spatial, and temporal expression of two sea urchin metallothionein genes
A:Reference number: A38739; MUID:91177920
A:Accession: A38739
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-65 <NEM>
A:Cross-references: GB:M59822; NID:g161535; PIDN:AAA30062.1; PID:g161536
!Superfamily: metallothionein

Query Match	40.5%;	Score 53;	DB 2;	Length 65;
Best Local Similarity	48.0%;	Pred. No. 11;		
Matches 12;	Conservative 0;	Mismatches 11;	Indels 2;	Gaps 1;

```

QY      2 CTGAXDCT--SCTXACTGCGXCPNA 24
          |||  ||  |  |||  |||
DB     17 CTGQDCCTIGKCKDKDGTCCGKCSNA 41

```

RESULT 15

ferrodoxin-like protein (AF083948) [Imported] - Agrobacterium tumefaciens (strain C58, C58)
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
 C:Accession: G98188
 R:Goodner, B.; Hillke, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A:Reference number: A97359; PMID:11743194
 A:Accession: G98188
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-166 <RUR>

A,Cross-references: GB:AE007870; PIDs:AAK89033.1; PID:915158827; GSDB:GNO017070
C,Genetics:
A,Gene: AGR_L_920
A,Map position: linear chromosome

Query Match	40.5%;	Score 53;	DB 2;	Length 166;
Best Local Similarity	26.2%;	Pred. No. 19;		
Matches	11;	Conservative	1;	Mismatches 6;
				Indels 24;
				Gaps 1

```
QY      4 GADACTSC-----TAACTGCGXC 21
          |   || :|
Db     104 GNVDCQACRDCAPTEARPRRGGPFVPAALVEDACTGCGAC 145
```

Search completed: October 21, 2002, 16:40:09
Job time : 6.76471 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:23 : Search time 2.82353 Seconds
(without alignments)
329.116 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XGTGXADCTSCFXACTGCGXCPNA 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	52.7	1176	1	NIR_NEUCR
2	65	49.6	1104	1	NIR_EMENT
3	60	45.8	325	1	Y856_TREPA
4	57	43.5	63	1	M84A_DROME
5	55.5	42.4	79	1	MT2_MALDO
6	53	40.5	25	1	MT_AGABT
7	53	40.5	65	1	MTB_STRPU
8	52	39.7	60	1	MTB_CHIHA
9	52	39.7	60	1	MTB_CHAAC
10	52	39.7	60	1	MTB_PAGBE
11	52	39.7	60	1	MTB_LIZAV
12	52	39.7	60	1	LRP2_RAT
13	51.5	39.3	60	1	MTA_SPAAU
14	51.5	39.3	72	1	MT2A_LYCES
15	51.5	39.3	73	1	MT2Y_LYCES
16	51.5	39.3	74	1	MT2_NICPL
17	51.5	39.3	78	1	MT2_ACTCH
18	51.5	39.3	178	1	CHHC_BOOMO
19	51.5	39.3	194	1	KRUB_HUMAN
20	51.5	39.3	1798	1	LMB2_HUMAN
21	51.5	39.3	2215	1	MTB_ONCMY
22	51	38.9	60	1	VGLM_HANTH
23	51	38.9	1134	1	MT_AMBME
24	50.5	38.5	60	1	MT_ESOUL
25	50.5	38.5	60	1	MT2A_RABIT
26	50.5	38.5	62	1	MT2C_RABIT
27	50.5	38.5	68	1	MT3_HORSE
28	50.5	38.5	68	1	MT3_HUMAN
29	50.5	38.5	68	1	MT3_PIG
30	50.5	38.5	72	1	MT2X_LYCES
31	50.5	38.5	713	1	TS4A_GALTA
32	50.5	38.5	713	1	YC81_CAREL
33	50.5	38.5	1291	1	Q19981

34	50.5	38.5	1592	1	SORL_CHICK	Q98930 g sortilin-
35	50.5	38.5	1799	1	LMB2_MOUSE	O61292 mus musculu
36	50.5	38.5	1801	1	LMB2_MOUSE	P15800 rattus norv
37	50	38.2	60	1	MTB_DICLA	O9PL99 dicentrarch
38	50	38.2	60	1	MTB_SALSA	P52720 salmo salar
39	50	38.2	60	1	MT_PERFL	P52725 perca fluvi
40	50	38.2	60	1	MT_ZOAVI	P52728 zoares viv
41	50	38.2	74	1	M84B_DROME	O01643 drosophila
42	50	38.2	279	1	NAPG_HAEIN	P44652 haemophilus
43	49.5	37.8	34	1	SCXM_SCOMA	P80719 scorpiu mau
44	49.5	37.8	60	1	MT2_CYPCA	O91910 cypripus ca
45	49.5	37.8	60	1	MT_BRAPE	P52722 brachydania

ALIGNMENTS

RESULT 1	ID	NIR_NEUCR	STANDARD:	PRT: 1176 AA.
AC	P38681:			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Nitrite reductase [NAD(P)H] (EC 1.6.6.4).			
GN	NIT-6.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_Taxid=5141;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-74-OR23-1A:			
RX	MEDLINE-93224461; PubMed-8096840;			
RA	Exley G.E., Colandene J.D., Garrett R.H.:			
RT	"Molecular cloning, characterization, and nucleotide sequence of			
RT	nit-6, the structural gene for nitrite reductase in Neurospora			
RT	crassa.";			
RT	J. Bacteriol. 175:2379-2392(1993).			
CC	-1- CATALYTIC ACTIVITY: 3 NAD(P)H + nitrite -> 3 NAD(P)(+) + NH(4)OH + H(2)O.			
CC	-1- COFACTOR: THIS ENZYME IS A FAD FLAVOPROTEIN THAT ALSO CONTAINS			
CC	A SIROHEME AND ONE 2FE-2S IRON-SULFUR CENTER.			
CC	-1- PATHWAY: SECOND STEP IN NITRATE ASSIMILATION (DENITRIFICATION).			
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC	-1- INDUCTION: BY NITRATE.			
CC	-1- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN			
CC	FOUND IN NITRATE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND			
CC	SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/			
CC	OR SEND AN EMAIL TO license@sib-sib.ch).			
CC	-----			
DR	EMBL: L07391; -; NOT_ANNOTATED_CDS.			
DR	PIR: A49848; A49848.			
DR	InterPro: IPR001327; FAD_Pyr_redox.			
DR	InterPro: IPR000660; Nit_Sir.			
DR	InterPro: IPR001281; Rieske.			
DR	Pfam: PR01070; NIT_SIR; 1.			
DR	Pfam: PR00070; Pyr_redox; 1.			
DR	Pfam: PF00355; Rieske; 1.			
DR	PRINTS: PR00397; SIROHAEM.			
DR	PROSITE: PS00365; NIT_SIR; 1.			
KW	Oxidoreductase; FAD; Flavoprotein; Iron-sulfur; Nitrate assimilation;			
KW	Heme; NADP.			
FT	NP_BIND 26 60 FAD (POTENTIAL);			
FT	NP_BIND 183 215 NAD(P)H (POTENTIAL);			
FT	METAL 717 717 IRON-SULFUR (2FE-2S) (POTENTIAL).			

```

FT METAL 723 723 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 757 757 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 761 761 IRON-SULFUR (2FE-2S) AND SIROHEME
          (BY SIMILARITY).
FT DOMAIN 998 1054 PRO/SER-RICH.
SQ SEQUENCE 1176 AA; 127367 MW; FFC7DCE6E6F80C710 CRC64;

Query Match 52.7%; Score 69; DB 1; Length 1176;
Best Local Similarity 60.0%; Pred. No. 0.072;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXC 21
   1 1 1 1 1 1 1 1
Db 515 CTSLGDLKCTKAGTGGCGC 534

RESULT 2
NIR_EMENI STANDARD; PRT; 1104 AA.
AC P22944;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Nitrite reductase [NAD(P)H] (EC 1.6.6.4).
GN NITR.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90382664; PubMed=2205530;
RA Johnstone I.L., McCabe P.C., Greaves P., Gurr S.J., Cole G.E.,
RA Brown A.D., Umles S.E., Clutterbuck A.J., Kinghorn J.R., Innis M.A.;
RA "Isolation and characterisation of the crnA-nla-nlaD gene cluster
RA for nitrate assimilation in Aspergillus nidulans.";
RL Gene 90:181-192(1990).
CC -1 CATALYTIC ACTIVITY: 3 NAD(P)H + nitrite = 3 NAD(P)(+) + NH(4)OH +
CC H(2)O.
CC -1 COFACTOR: THIS ENZYME IS A FAD FLAVOPROTEIN THAT ALSO CONTAINS
CC A SIROHEME AND ONE 2FE-2S IRON-SULFUR CENTER.
CC -1 PATHWAY: SECOND STEP IN NITRATE ASSIMILATION (DENITRIFICATION).
CC -1 SUBUNIT: HOMODIMER.
CC -1 SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58289; AAA33315.1; -.
DR PIR; JH0181; JH0181.
DR InterPro; IPR001327; FAD_DYF_redox.
DR InterPro; IPR000660; NIT_Sir.
DR Pfam; PF01077; NIT_SIR; 1.
DR Pfam; PF00070; PYR_redox; 1.
DR Pfam; PF00355; Rlseske; 1.
DR PRINTS; PR00397; SIROHEM.
DR PROSITE; PS00365; NIT_SIR; 1.
KW Oxidoreductase; FAD; Flavoprotein; Iron-sulfur; Nitrate assimilation;
KW Heme; NADP.
FT NP_BIND 44 79 FAD (POTENTIAL).
FT NP_BIND 146 176 NAD(P)H (POTENTIAL).
FT METAL 720 720 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 726 726 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 760 760 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 764 764 IRON-SULFUR (2FE-2S) AND SIROHEME

```

```

FT SQ SEQUENCE 1104 AA; 122731 MW; E5D00436E1E86304 CRC64;
          (BY SIMILARITY).
Query Match 49.6%; Score 65; DB 1; Length 1104;
Best Local Similarity 60.0%; Pred. No. 0.22;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXC 21
   1 1 1 1 1 1 1 1
Db 519 CKTIADVKCTKAGTGGCGC 538

RESULT 3
Y856_TREPA
ID Y856_TREPA STANDARD; PRT; 325 AA.
AC O83828;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0856 precursor.
GN TP0856.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NICHOLS.
RX MEDLINE=9832770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Venter B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Hatcher J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RA spirochete.";
RL Science 281:375-388(1998).
CC -1 SIMILARITY: BELONGS TO THE UPF0164 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001256; AAC65828.1; -.
DR TIGR; TP0856; -.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 325 HYPOTHETICAL PROTEIN TP0856.
SQ SEQUENCE 325 AA; 34054 MW; F9CFDCBD253C07D2 CRC64;

Query Match 45.8%; Score 60; DB 1; Length 325;
Best Local Similarity 47.6%; Pred. No. 0.38;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 3 TGXADCTCTACTGCGXCPN 23
   1 1 1 1 1 1 1 1
Db 233 TGGQCKTKTSGRCGTGTCNCN 253

RESULT 4
M84A_DROME
ID M84A_DROME STANDARD; PRT; 63 AA.
AC G01642; G9VIA3;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Male specific sperm protein Mst84A.
GN Mst84DA OR CGI7946.

```

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=92102953; PubMed=1684716;
 RA Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
 RT "A cluster of four genes selectively expressed in the male germ line
 of Drosophila melanogaster.";
 RL Mech. Dev. 35:143-151(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Belt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jatalil B.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
 RA Palazotto M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA She B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- TISSUE SPECIFICITY: TESTIS.
 CC -1- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
 CC -1- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
 MOTIFS.
 CC -1- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: X67703; CAA47937.1; -
 DR EMBL: AE003672; AAP54022.1; -
 DR HSSP: P01180; INPO.

DR FlyBase: FBgn0004172; Mst84Da.
 KW Spermatogenesis; Repeat; Multigene family.
 SQ SEQUENCE 63 AA; 5806 MW; BF84CD74CBCEFD3F CRC64;
 Query Match 43.5%; Score 57; DB 1; Length 63;
 Best Local Similarity 45.0%; Pred. No. 0.27; Mismatches 0; Gaps 0;
 Matches 9; Conservative 0; Indels 11; Indels 0;
 QY 2 CTGXADCTCTXACTGCGXC 21
 Db 20 CGGCGCGCGCGCGCGCGC 39
 RESULT 5
 WT2_MALDO STANDARD; PRT; 79 AA.
 ID WT2_MALDO
 AC 024058;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Metallothionein-like protein type 2.
 GN MT1.
 OS Malus domestica (Apple) (Malus sylvestris).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Rosales; Rosaceae; Maloideae; Malus.
 NC NCB1_TaxID=3750;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fruit cortical tissue;
 RA Reid S.J., Ross G.S.;
 RT "Up-regulation of two cDNA clones encoding metallothionein-like
 proteins in apple fruit during cool storage.";
 RL Physiol. Plantarum 100:183-189(1997).
 CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 RESIDUES THAT BIND VARIOUS HEAVY METALS.
 CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: U61973; AAC23697.1; -
 DR InterPro: IPR000347; Metallothion_15.
 DR Pfam: PF01439; Metallothio_2; 1.
 DR ProDom: PD001611; Metallothion.15; 1.
 KW Metal-binding; Metal-thiolate cluster; Chelation.
 SQ SEQUENCE 79 AA; 7836 MW; 8ADC58D18B644CC CRC64;
 Query Match 42.4%; Score 55.5; DB 1; Length 79;
 Best Local Similarity 40.0%; Pred. No. 0.49;
 Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;
 QY 2 CTGXADC---TSCTXACTGCGXCPN 23
 Db 5 CGGKCGGCGVSCGSGCGCGMAPD 29
 RESULT 6
 MT_AGABI
 ID MT_AGABI STANDARD; PRT; 25 AA.
 AC F04358;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Metallothionein (MT).
 OS Agaricus bisporus (Common mushroom).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Agaricaceae; Agaricus.

```

OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE.
RC STRAIN-A-32; TISSUE=Mycellium;
RA Muenger K., Lerch K.;
RT "Copper metallothionein from the fungus Agaricus bisporus: chemical
RL and spectroscopic properties.";
RL Biochemistry 24:6751-6756(1985).
CC -1- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR
CC SEQUESTRATION OF TOXIC METAL IONS. BINDS SIX COPPER (COPROUS)
CC IONS.
CC -1- MISCELLANEOUS: THE ABSORPTION, LUMINESCENT, AND STEREOPTICAL
CC PROPERTIES OF THE COPPER MT ARE ATTRIBUTED TO THE METAL-THIOLATE
CC COMPLEX BECAUSE THEY ARE NOT PRESENT IN THE APOPROTEIN.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 8.
DR PIR: A03286; SMMR.
DR KW Metal-binding; Metal-thiolate cluster; Chelation; Copper.
SQ SEQUENCE 25 AA; 2233 MW; 33AB70F21023CEFF7 CRC64;

OY 2 CTGXADCT----SCTXACTGCG 19
OY 1:1 11 11 1:111
OY 5 CSGASCTCAGSGCT--CSGCC 24

Db

RESULT 7
MTB_STRPU
ID MTB_STRPU STANDARD: PRT: 65 AA.
AC 027287;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Metallothionein-B (MTB).
GN MTB1 OR MTB.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OX Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87172743; PubMed=3561398;
RA Wilkinson D.G., Nemer M.;
RT "Metallothionein genes MTA and MTB expressed under distinct
RT quantitative and tissue-specific regulation in sea urchin embryos.";
RL Mol. Cell. Biol. 7:48-58(1987).
RN [2]
RP SEQUENCE FROM N.A.
RP MEDLINE=9117920; PubMed=2007604;
RA Nemer M., Thornton R.D., Stuebing E.W., Harlow P.;
RT "Structure, spatial, and temporal expression of two sea urchin
RT metallothionein genes, SPMTB1 and SPMTA.";
RL J. Biol. Chem. 266:6586-6593(1991).
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: M15989; AAA30063.1; -
DR EMBL: M59822; AAA30062.1; -.
DR HSSP: P04734; 10LI.
DR InterPro: IPR001396; Metallothion_Echnd.

```

```

DR PRINTS:PR00873; MTECHINOIDEA.
DR Metal-binding; Metal-thiolate cluster; Chelation.
SQ SEQUENCE 65 AA; 6470 MW; 12E1C08064D9121A5 CRC64;

Query Match 40.5%; Score 53; DB 1; Length 65;
Best Local Similarity 48.0%; Pred. NO. 0.87;
Matches 12; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

Oy 2 CTGXADCT--SCTXACTGCGXCPMA 24
||| ||| | | ||| |||
Db 17 CTGDCCTTGKCKCKDCTCGKCSNA 41

RESULT 8
MTB_CHAAC
AC MTB_CHAAC STANDARD: PRT: 60 AA.
ID P52724;
DI 01-OCT-1996 (Rel. 34, Created)
DI 01-OCT-1996 (Rel. 34, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS Chenocephalus aceratus (White crocodile fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorphi; Perciformes;
OC Noronthenioidae; Chanichthyidae; Chanocephalus.
OX NCBI_TaxId=36190;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97233027; PubMed=9078263;
RA Scudiero R., Carignale V., Riggio M., Capasso C., Capasso A.,
RA Kille P., di Prisco G., Parisi E.;
RT "Difference in hepatic metallothionein content in Antarctic
RT red-blooded and haemoglobinless fish: undetectable metallothionein
RT levels in haemoglobinless fish is accompanied by accumulation of
RT untranslated metallothionein mRNA.";
RL Biochem. J. 322:207-211(1997).

-1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
-1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
-1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on lists
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: 272483; CAA96564.1; -.
DR HSSP: P02802; IDRS.
DR InterPro: IPR003019; Metallthion.
DR Pfam: PF00131; metalthio; 1.
DR PRINTS: PR00860; MTVERTBRATE.
DR PROSITE: PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster; Chelation.
FT DOMAIN 1 28 BETA.
FT DOMAIN 29 60 ALPHA.
FT METAL 4 4 CLUSTER B.
FT METAL 6 6 CLUSTER B.
FT METAL 12 12 CLUSTER B.
FT METAL 14 14 CLUSTER B.
FT METAL 18 18 CLUSTER B.
FT METAL 20 20 CLUSTER B.
FT METAL 23 23 CLUSTER B.

```



```

FT METAL 25 25 CLUSTER B.
FT METAL 28 28 CLUSTER B.
FT METAL 32 32 CLUSTER A.
FT METAL 33 33 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 36 36 CLUSTER A.
FT METAL 40 40 CLUSTER A.
FT METAL 43 43 CLUSTER A.
FT METAL 47 47 CLUSTER A.
FT METAL 49 49 CLUSTER A.
FT METAL 54 54 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA; 6006 MW; EF61F4AD61B9124A CRC64;

Query Match 39.7%; Score 52; DB 1; Length 60;
Best Local Similarity 43.5%; Pred. No. 1.1;
Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

QY 2 CTGXADCTCTCTACTGC--GXCP 22
Db 14 CGGSCCTCNC--SCTSCCKSCCP 34

RESULT 9
MTB_CHIHA STANDARD; PRT; 60 AA.
ID 013259;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS Chionodraco hamatus (Antarctic teleost icefish),
OS Morone saxatilis (Striped bass), and
OS Gymnodraco acuticeps (Antarctic dragonfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Channichthyidae; Chionodraco.
OX NCBI_TaxID=36188, 34816, 8218;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=C.hamatus; TISSUE=Liver;
RX MEDLINE=98264843; PubMed=9601077;
RA Cargiale V., Scudiero R., Capasso C., Capasso A., Kille P.,
RA di Prisco G., Parisi E.;
RT "Cadmium-induced differential accumulation of metallothionein isoforms
RT in the Antarctic icefish, which exhibits no basal metallothionein
RT protein but high endogenous mRNA levels.";
RL Biochem. J. 332:475-481(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.saxatilis;
RA Leclerc G.M., Leclerc G.J., Ely B.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=G.acuticeps; TISSUE=Liver;
RX MEDLINE=99334610; PubMed=10406107;
RA Bargelloni L., Scudiero R., Parisi E., Cargiale V., Capasso C.,
RA Patarnello T.;
RT "Metallothioneins in antarctic fish: evidence for independent
RT duplication and gene conversion.";
RL Mol. Biol. Evol. 16:885-887(1999).
RN [4]
RP FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
RN RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
RN -1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
RN FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
RN DOMAIN AND ARE COORDINATED VIA CYSTEINE, THIOLATE BRIDGES TO 11
RN CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
RN BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
RN -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: Y12581; CAA73160.1; -.
DR EMBL: AF091100; AAC62501.1; -.
DR EMBL: AJ007561; CAA07556.1; -.
DR HSSP: P02802; IDES.
DR InterPro: IPR003019; Metallothion.
DR InterPro: IPR000006; Metallothion_vert.
DR Pfam: PF00131; metalthio.1.
DR PRINTS: PR00860; MIVERTERRATE.
DR PROSITE: PS00203; METALLOTHIONEIN_VRT; 1.
DR Metal-binding; Metal-thiolate cluster; Chelation.
KW DOMAIN 1 28
FT METAL 29 60 ALPHA.
FT METAL 4 4 CLUSTER B (BY SIMILARITY).
FT METAL 6 6 CLUSTER B (BY SIMILARITY).
FT METAL 12 12 CLUSTER B (BY SIMILARITY).
FT METAL 14 14 CLUSTER B (BY SIMILARITY).
FT METAL 18 18 CLUSTER B (BY SIMILARITY).
FT METAL 20 20 CLUSTER B (BY SIMILARITY).
FT METAL 23 23 CLUSTER B (BY SIMILARITY).
FT METAL 25 25 CLUSTER B (BY SIMILARITY).
FT METAL 28 28 CLUSTER B (BY SIMILARITY).
FT METAL 32 32 CLUSTER A (BY SIMILARITY).
FT METAL 33 33 CLUSTER A (BY SIMILARITY).
FT METAL 35 35 CLUSTER A (BY SIMILARITY).
FT METAL 36 36 CLUSTER A (BY SIMILARITY).
FT METAL 40 40 CLUSTER A (BY SIMILARITY).
FT METAL 43 43 CLUSTER A (BY SIMILARITY).
FT METAL 47 47 CLUSTER A (BY SIMILARITY).
FT METAL 49 49 CLUSTER A (BY SIMILARITY).
FT METAL 54 54 CLUSTER A (BY SIMILARITY).
FT METAL 58 58 CLUSTER A (BY SIMILARITY).
FT METAL 59 59 CLUSTER A (BY SIMILARITY).
SQ SEQUENCE 60 AA; 5978 MW; 5966F5AC61BC424B CRC64;

Query Match 39.7%; Score 52; DB 1; Length 60;
Best Local Similarity 43.5%; Pred. No. 1.1;
Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

QY 2 CTGXADCTCTCTACTGC--GXCP 22
Db 14 CGGSCCTCNC--SCTSCCKSCCP 34

RESULT 10
MTB_PAGBE STANDARD; PRT; 60 AA.
ID 092145;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS Pagohenia bernacchii (Emerald rockcod) (Trematomus bernacchii),
OS Chionodraco rastrospinosus (Ocellated icefish),
OS Notothenia coriiceps neglecta (Black rockcod) (Yellowbelly rockcod),
OS Pagohenia borhgrevinkii (Bald rockcod) (Trematomus borhgrevinkii),
OS and Paracheanichthys charcoti.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Notothenidae; Trematomus.
OX NCBI_TaxID=40690, 34790, 8209, 8213, 36187;
RN [1]
RP SEQUENCE FROM N.A.

```

RC SPECIES=P.bernacchi, and C.astrospinosus; TISSUE=Liver;
 RA MEDLINE=97233027; PubMed=9078263;
 RX Scudiero R., Carginale V., Riggio M., Capasso C., Capasso A.,
 RA Kille P., di Prisco G., Parisi E.,
 RT "Difference in hepatic metallothionein content in Antarctic
 RT red-blooded and haemoglobinless fish: undetectable metallothionein
 RT levels in haemoglobinless fish is accompanied by accumulation of
 RT untranslated metallothionein mRNA.";
 RL Biochem. J. 322:207-211(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=N.c. neglecta, P.borchgrevinkii, and P. charcotii;
 RC TISSUE=Liver;
 RA MEDLINE=99334610; PubMed=10406107;
 RX Bargelloni L., Scudiero R., Parisi E., Carginale V., Capasso C.,
 RA Patarnello T.,
 RT "Metallothioneins in antarctic fish: evidence for independent
 RT duplication and gene conversion".
 RL Mol. Biol. Evol. 16:885-897(1999).
 CC -I- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -I- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -I- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z72485; CAA96566.1; -;
 DR EMBL; Z72484; CAA96565.1; -;
 DR EMBL; AJ006485; CAA07064.1; -;
 DR EMBL; AJ007563; CAA07558.1; -;
 DR EMBL; AJ007951; CAA07786.1; -;
 DR HSSP; P02802; IDPS.
 DR InterPro; IPR003019; Metallthion.
 DR InterPro; IPR000006; Metallthion_vert.
 DR Pfam; PF00131; metalchio; 1.
 DR PRINTS; PR00860; MTVERTEBRATE.
 DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
 KW Metal-binding; Metal-thiolate cluster; Chelation.
 FT DOMAIN 1 28 BETA.
 FT METAL 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B.
 FT METAL 6 6 CLUSTER B.
 FT METAL 12 12 CLUSTER B.
 FT METAL 14 14 CLUSTER B.
 FT METAL 18 18 CLUSTER B.
 FT METAL 20 20 CLUSTER B.
 FT METAL 23 23 CLUSTER B.
 FT METAL 25 25 CLUSTER B.
 FT METAL 28 28 CLUSTER B.
 FT METAL 32 32 CLUSTER A.
 FT METAL 33 33 CLUSTER A.
 FT METAL 35 35 CLUSTER A.
 FT METAL 36 36 CLUSTER A.
 FT METAL 40 40 CLUSTER A.
 FT METAL 43 43 CLUSTER A.
 FT METAL 47 47 CLUSTER A.
 FT METAL 49 49 CLUSTER A.
 FT METAL 54 54 CLUSTER A.
 FT METAL 58 58 CLUSTER A.
 FT METAL 59 59 CLUSTER A.
 SQ SEQUENCE 60 AA; 5992 MW; E866F4AD61BC424A CRC64;

Query Match 39.7%; Score 52; DB 1; Length 60;

Best Local Similarity 43.5%; Pred. No. 1.1;
 Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;
 QY 2 CTCXADCTSTXACTGC--GXCP 22
 DB 14 CGSGCTCTNC--SCTSCKRSCCP 34
 RESULT 11
 MT_LIZAU
 ID MT_LIZAU STANDARD; PRT; 60 AA.
 AC O13257;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Metallothionein (MT).
 GN MT.
 OS Liza aurata (Golden grey mullet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Mugilomorpha; Mugilidae;
 OC Liza
 OX NCBI_TaxID=48191;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Cousinou M., Lopez-Barea J., Dorado G.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -I- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -I- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U93207; AAB51591.1; -;
 DR HSSP; P02802; IDPS.
 DR InterPro; IPR003019; Metallthion.
 DR InterPro; IPR000006; Metallthion_vert.
 DR Pfam; PF00131; metalchio; 1.
 DR PRINTS; PR00860; MTVERTEBRATE.
 DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
 KW Metal-binding; Metal-thiolate cluster; Chelation.
 FT DOMAIN 1 28 BETA.
 FT METAL 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B.
 FT METAL 6 6 CLUSTER B.
 FT METAL 12 12 CLUSTER B.
 FT METAL 14 14 CLUSTER B.
 FT METAL 18 18 CLUSTER B.
 FT METAL 20 20 CLUSTER B.
 FT METAL 23 23 CLUSTER B.
 FT METAL 25 25 CLUSTER B.
 FT METAL 28 28 CLUSTER B.
 FT METAL 32 32 CLUSTER A.
 FT METAL 33 33 CLUSTER A.
 FT METAL 35 35 CLUSTER A.
 FT METAL 36 36 CLUSTER A.
 FT METAL 40 40 CLUSTER A.
 FT METAL 43 43 CLUSTER A.
 FT METAL 47 47 CLUSTER A.
 FT METAL 49 49 CLUSTER A.
 FT METAL 54 54 CLUSTER A.

```

FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SO SEQUENCE 60 AA; 5992 MW; 3EAA86F093C4701 CRC64;

Query Match 39.7%; Score 52; DB 1; Length 60;
Best Local Similarity 43.5%; Pred. No. 1.1;
Matches 10; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

OY 2 CTGAXADCTSCXTACTGC--GXCP 22
Db 14 CGGSCCTCTNC--SCTSCKSCCP 34

RESULT 12
LRP2_RAT STANDARD; PRT; 4660 AA.
ID LRP2_RAT
AC P98138;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Low-density lipoprotein receptor-related protein 2 precursor (Megalin)
DE (Glycoprotein 330).
GN LRP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;
RX MEDLINE=95024033; PubMed=7937880;
RA Saito A., Pietromonaco S., Loo A.K.C., Fargnhar M.G.;
RT "Complete cloning and sequencing of rat gp330/megalin, a
RT distinctive member of the low density lipoprotein receptor gene
RT family.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9725-9729(1994).
RN [2]
RP FUNCTION.
RX MEDLINE=95386696; PubMed=7544804;
RA Moestup S.K., Cui S., Vorum H., Bregengaard C., Bjorn S.E.,
RA Norris K., Gilleman J., Christensen E.I.;
RT "Evidence that epithelial glycoprotein 330/megalin mediates uptake of
RT polybasic drugs.";
RL J. Clin. Invest. 96:1404-1413(1995).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=94172242; PubMed=7510321;
RA Zheng G., Bachlinsky D.R., Stamenkovic I., Strickland D.K., Brown D.,
RA Andres G., McCluskey R.T.;
RT "Organ distribution in rats of two members of the low-density
RT lipoprotein receptor gene family, gp330 and LRP/alpha 2MR, and the
RT receptor-associated protein (RAP).";
RL J. Histochem. Cytochem. 42:531-542(1994).
CC -1- FUNCTION: BINDS PLASMINOGEN, EXTRACELLULAR MATRIX COMPONENTS,
CC PLASMINOGEN ACTIVATOR-PLASMINOGEN ACTIVATOR INHIBITOR TYPE I
CC COMPLEX, APOLIPOPROTEIN E-ENRICHED BETA-VLDL, LIPOPROTEIN LIPASE,
CC LACTOFERRIN, CLUSTERIN AND CALCIUM.
CC -1- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS
CC APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B.
CC -1- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-
CC ASSOCIATED PROTEIN (RAP).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXPRESSED IN
CC CLATHRIN-COATED PITS: A SOLUBLE FORM IS POSSIBLY DERIVED BY
CC CLEAVAGE AT THE CELL SURFACE.
CC -1- TISSUE SPECIFICITY: EPITHELIAL CELLS OF KIDNEY GLOMERULUS AND
CC PROXIMAL TUBULE, LUNG, EPIDIDYMS, YOLK SAC, AMONG OTHERS.
CC -1- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS.
CC -1- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; I34049; AAA51369.1; -.
DR HSP; O07954; ICR8.
DR GlycosultedB; P98158; -.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000033; LDL_receptor_rep.
DR Pfam; PF00057; ldl_recept_a; 36.
DR Pfam; PF00058; ldl_recept_b; 33.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF_LIKE; 15.
DR SMART; SM00192; LDLR; 36.
DR SMART; SM00135; LY; 35.
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01209; LDLR_1; 31.
DR PROSITE; PS00068; LDLR_2; 36.
KW Glycoprotein; Repeat; Endocytosis; Coated pits; Transmembrane;
KW Receptor; EGF-like domain; SH3-binding; Signal.
FT SIGNAL 1 25
FT CHAIN 26 4660
FT FT 26 4425
FT FT 4426 4446
FT FT 4447 4660
FT DOMAIN 26 64
FT DOMAIN 65 105
FT DOMAIN 106 144
FT DOMAIN 145 181
FT DOMAIN 181 219
FT DOMAIN 220 258
FT DOMAIN 258 307
FT DOMAIN 308 346
FT DOMAIN 347 385
FT DOMAIN 385 435
FT DOMAIN 435 476
FT DOMAIN 476 519
FT DOMAIN 521 566
FT DOMAIN 568 611
FT DOMAIN 612 652
FT DOMAIN 658 704
FT DOMAIN 704 793
FT DOMAIN 793 835
FT DOMAIN 837 879
FT DOMAIN 881 923
FT DOMAIN 923 1013
FT DOMAIN 1013 1061
FT DOMAIN 1064 1103
FT DOMAIN 1103 1146
FT DOMAIN 1146 1186
FT DOMAIN 1186 1225
FT DOMAIN 1225 1269
FT DOMAIN 1270 1308
FT DOMAIN 1311 1351
FT DOMAIN 1350 1390
FT DOMAIN 1391 1430
FT DOMAIN 1430 1479
FT DOMAIN 1479 1520
FT DOMAIN 1520 1563
FT DOMAIN 1567 1609
FT DOMAIN 1611 1654
FT DOMAIN 1656 1696
FT DOMAIN 1701 1742
FT DOMAIN 1791 1832
FT DOMAIN 1834 1882
FT DOMAIN 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 923
FT FT 923 1013
FT FT 1013 1061
FT FT 1064 1103
FT FT 1103 1146
FT FT 1146 1186
FT FT 1186 1225
FT FT 1225 1269
FT FT 1270 1308
FT FT 1311 1351
FT FT 1350 1390
FT FT 1391 1430
FT FT 1430 1479
FT FT 1479 1520
FT FT 1520 1563
FT FT 1567 1609
FT FT 1611 1654
FT FT 1656 1696
FT FT 1701 1742
FT FT 1791 1832
FT FT 1834 1882
FT FT 1884 1930
FT FT 4425 4446
FT FT 4447 4660
FT FT 64 105
FT FT 106 144
FT FT 145 181
FT FT 181 219
FT FT 220 258
FT FT 258 307
FT FT 308 346
FT FT 347 385
FT FT 385 435
FT FT 435 476
FT FT 476 519
FT FT 521 566
FT FT 568 611
FT FT 612 652
FT FT 658 704
FT FT 704 793
FT FT 793 835
FT FT 837 879
FT FT 881 9
```

```

FT DOMAIN 1932 1972 LDL-RECEPTOR CLASS B 18.
FT DOMAIN 1973 2013 LDL-RECEPTOR CLASS B 19.
FT DOMAIN 2019 2060 EGF-LIKE 8.
FT DOMAIN 2108 2156 LDL-RECEPTOR CLASS B 20.
FT DOMAIN 2158 2201 LDL-RECEPTOR CLASS B 21.
FT DOMAIN 2203 2245 LDL-RECEPTOR CLASS B 22.
FT DOMAIN 2247 2289 LDL-RECEPTOR CLASS B 23.
FT DOMAIN 2291 2332 LDL-RECEPTOR CLASS B 24.
FT DOMAIN 2343 2384 EGF-LIKE 9.
FT DOMAIN 2432 2477 LDL-RECEPTOR CLASS B 25.
FT DOMAIN 2479 2518 LDL-RECEPTOR CLASS B 26.
FT DOMAIN 2520 2562 LDL-RECEPTOR CLASS B 27.
FT DOMAIN 2564 2604 LDL-RECEPTOR CLASS B 28.
FT DOMAIN 2605 2647 LDL-RECEPTOR CLASS B 29.
FT DOMAIN 2652 2694 EGF-LIKE 10.
FT DOMAIN 2699 2739 LDL-RECEPTOR CLASS A 16.
FT DOMAIN 2740 2778 LDL-RECEPTOR CLASS A 17.
FT DOMAIN 2779 2820 LDL-RECEPTOR CLASS A 18.
FT DOMAIN 2821 2862 LDL-RECEPTOR CLASS A 19.
FT DOMAIN 2863 2903 LDL-RECEPTOR CLASS A 20.
FT DOMAIN 2906 2947 LDL-RECEPTOR CLASS A 21.
FT DOMAIN 2948 2992 LDL-RECEPTOR CLASS A 22.
FT DOMAIN 2993 3031 LDL-RECEPTOR CLASS A 23.
FT DOMAIN 3032 3072 LDL-RECEPTOR CLASS A 24.
FT DOMAIN 3073 3112 LDL-RECEPTOR CLASS A 25.
FT DOMAIN 3113 EGF-LIKE 11.
FT DOMAIN 3154 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3241 3282 LDL-RECEPTOR CLASS B 30.
FT DOMAIN 3284 3333 LDL-RECEPTOR CLASS B 31.
FT DOMAIN 3335 3377 LDL-RECEPTOR CLASS B 32.
FT DOMAIN 3379 3420 LDL-RECEPTOR CLASS B 33.
FT DOMAIN 3421 3461 LDL-RECEPTOR CLASS B 34.
FT DOMAIN 3467 3511 EGF-LIKE 13.
FT DOMAIN 3512 3552 LDL-RECEPTOR CLASS A 26.
FT DOMAIN 3553 3593 LDL-RECEPTOR CLASS A 27.
FT DOMAIN 3594 3634 LDL-RECEPTOR CLASS A 28.
FT DOMAIN 3635 3675 LDL-RECEPTOR CLASS A 29.
FT DOMAIN 3678 3718 LDL-RECEPTOR CLASS A 30.
FT DOMAIN 3719 3758 LDL-RECEPTOR CLASS A 31.
FT DOMAIN 3759 3797 LDL-RECEPTOR CLASS A 32.
FT DOMAIN 3798 3836 LDL-RECEPTOR CLASS A 33.
FT DOMAIN 3842 3882 LDL-RECEPTOR CLASS A 34.
FT DOMAIN 3883 3924 LDL-RECEPTOR CLASS A 35.
FT DOMAIN 3928 3966 LDL-RECEPTOR CLASS A 36.
FT DOMAIN 3968 4008 EGF-LIKE 14.
FT DOMAIN 4009 4050 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4156 4197 LDL-RECEPTOR CLASS B 35.
FT DOMAIN 4199 4241 LDL-RECEPTOR CLASS B 36.
FT DOMAIN 4244 4284 LDL-RECEPTOR CLASS B 37.
FT DOMAIN 4332 4370 EGF-LIKE 16.
FT DOMAIN 4379 4413 EGF-LIKE 17.
FT SITE 4457 4463 SH3-BINDING (POTENTIAL).
FT SITE 4454 4460 SH3-BINDING (POTENTIAL).
FT SITE 4457 4463 SH2-BINDING (POTENTIAL).
FT SITE 4619 4625 SH2-BINDING (POTENTIAL).
FT SITE 4624 4630 SH3-BINDING (POTENTIAL).
FT SITE 1743 1745 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 4522 4527 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT SITE 4601 4606 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT DISULFID 28 40 BY SIMILARITY.
FT DISULFID 35 53 BY SIMILARITY.
FT DISULFID 47 62 BY SIMILARITY.
FT DISULFID 67 80 BY SIMILARITY.
FT DISULFID 74 93 BY SIMILARITY.
FT DISULFID 87 103 BY SIMILARITY.
FT DISULFID 108 120 BY SIMILARITY.
FT DISULFID 115 133 BY SIMILARITY.
FT DISULFID 127 142 BY SIMILARITY.
FT DISULFID 147 157 BY SIMILARITY.

```

Query Match 39.7%; Score 52; DB 1; Length 4660;
 Best Local Similarity 36.4%; Pred. No. 27;
 Matches 8; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

```

OY 2 CIGXADCTSCYACTGCGXCPN 23
Db 1320 CCGYSTCINLMSALCDGVFDCPN 1341

RESULT 13
MTA_SPAU
ID MTA_SPAU STANDARD; PRT: 60 AA.
AC P52727;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE Metallothionein A (MT A).
GN MTA.
OS Sparus aurata (Gilthead sea bream).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Sparidae; Sparus.
OX NCBITaxId=8175;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kille P., Olsson P.E.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Tom M.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Cousinou M., Lopez-Barea J., Dorado G.;
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -i- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -i- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X97276; CA65931.1; -
CC EMBL: U58774; AAC32738.1; -
CC EMBL: U93206; AAB51590.1; -
CC HSSP: P02802; IDRS.
CC InterPro: IPR003019; Metallothion.
CC InterPro: IPR000006; Metallothion_vert.
CC Pfam: PF00131; metalthio.1.
CC PRINTS: PR00860; MYVERTEBRATE.
CC PROSITE: PS00203; METALLOTHIONEIN_VRT; 1.
CC Metal-binding; Metal-thiolate cluster; Chelation.
FT DOMAIN 1 28 ALPHA.
FT METAL 4 4 CLUSTER B.
FT METAL 6 6 CLUSTER B.
FT METAL 12 12 CLUSTER B.
FT METAL 14 14 CLUSTER B.
FT METAL 18 18 CLUSTER B.
FT METAL 20 20 CLUSTER B.
FT METAL 23 23 CLUSTER B.
FT METAL 25 25 CLUSTER B.

```

```
FT METAL 28 28 CLUSTER B.
FT METAL 32 32 CLUSTER A.
FT METAL 33 33 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 36 36 CLUSTER A.
FT METAL 40 40 CLUSTER A.
FT METAL 43 43 CLUSTER A.
FT METAL 47 47 CLUSTER A.
FT METAL 49 49 CLUSTER A.
FT METAL 54 54 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA: 5966 MW: 6A66F79D02EC591B CRC64;

Query Match 39.3%; Score 51.5; DB 1; Length 60;
Best Local Similarity 47.6%; Pred. No. 1.3;
Matches 10; Conservative 1; Mismatches 7; Indels 3; Gaps 2;

OY 2 CTGADCTCTXACTGCGXCP 22
DB 20 CTN-CSCCKKSCSC--CP 37

RESULT 14
MT2A_LYCES STANDARD; PRT; 72 AA.
ID MT2A_LYCES
AC 040157;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Metallothionein-like protein type 2 A.
GN MTA.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, ATLAS CRAIG;
RA Whitelaw C.A., Lehuquet J.A., Thurman D.A., Tomsett A.B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L77963; AAB04674.1; -.
DR HSSP: P10969; IMGT.
DR InterPro: IPR000347; Metallothion_15.
DR Pfam: PF01439; Metallothio.2; 1.
DR ProDom: PD001611; Metallothion_15; 1.
KW Metal-binding; Metal-thiolate cluster; Chelation; Multigene family.
SQ SEQUENCE 72 AA: 7113 MW: 75FF89BB93AB1893 CRC64;

Query Match 39.3%; Score 51.5; DB 1; Length 72;
Best Local Similarity 40.0%; Pred. No. 1.4;
Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;

OY 2 CTGADCTCTXACTGCGXCPN 23
DB 4 CGGSCGCGSGCKGCGGCGGMPD 28

RESULT 15
MT2Y_LYCES
```

```
ID MT2Y_LYCES STANDARD; PRT; 73 AA.
AC 043513;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Metallothionein-like protein type 2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, BONNER BESTE; TISSUE=Root;
RA Grillich A., Herdik A., Balzer H., Stephan U., Baumann H.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z68310; CAA92652.1; -.
DR InterPro: IPR000347; Metallothion_15.
DR Pfam: PF01439; Metallothio.2; 1.
DR ProDom: PD001611; Metallothion_15; 1.
KW Metal-binding; Metal-thiolate cluster; Chelation; Multigene family.
SQ SEQUENCE 73 AA: 7129 MW: 01BAEB12B25F9E1 CRC64;

Query Match 39.3%; Score 51.5; DB 1; Length 73;
Best Local Similarity 40.0%; Pred. No. 1.5;
Matches 10; Conservative 1; Mismatches 11; Indels 3; Gaps 1;

OY 2 CTGADCTCTXACTGCGXCPN 23
DB 4 CGGSCGCGSGCKGCGGCGGMPD 28
```

Search completed: October 21, 2002, 16:38:45
Job time : 4.82353 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 21, 2002, 16:34:44 ; Search time 7.76471 Seconds
(without alignments)
534.712 Million cell updates/sec

Title: US-10-032-658-4

Perfect score: 131

Sequence: 1 XCTGXADCTCTACTGCGXCPNA 24

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	96.2	124	5	09U744 tenebrio mo
2	126	96.2	124	5	016121 tenebrio mo
3	126	96.2	148	5	016122 tenebrio mo
4	125	95.4	112	5	09U748 tenebrio mo
5	124	94.7	112	5	09U746 tenebrio mo
6	124	94.7	112	5	016119 tenebrio mo
7	124	94.7	112	5	016120 tenebrio mo
8	122	93.1	112	5	09U745 tenebrio mo
9	110	84.0	112	5	09U745 tenebrio mo
10	103	77.9	123	5	09NC09 dendroides
11	102	77.9	129	5	09NC09 dendroides
12	102	77.9	148	5	09NC08 dendroides
13	101	77.1	121	5	09NC08 dendroides
14	100	76.3	91	5	09NC03 dendroides
15	98	74.8	104	5	09NC03 dendroides
16	97	74.0	96	5	096419 dendroides

17	97	74.0	98	5	09NC06 dendroides
18	97	74.0	108	5	046351 dendroides
19	97	74.0	108	5	046352 dendroides
20	97	74.0	109	5	046346 dendroides
21	94	71.8	136	5	09NC02 dendroides
22	91	69.5	118	5	09NC07 dendroides
23	91	69.5	122	5	09NC04 dendroides
24	63	48.1	141	3	074998 dendroides
25	61	46.6	99	5	09N329 dendroides
26	59	45.0	871	2	09U2D2 dendroides
27	59	45.0	1044	3	09U9M3 dendroides
28	59	45.0	1044	3	000943 dendroides
29	58	44.3	189	11	09P527 dendroides
30	58	44.3	196	11	09P226 dendroides
31	58	44.3	719	4	060283 dendroides
32	56.5	43.1	2135	5	061077 dendroides
33	55.5	42.4	79	10	09UUX7 dendroides
34	55.5	42.4	134	5	024988 dendroides
35	55.5	42.4	234	5	09U064 dendroides
36	55.5	42.4	704	5	09U048 dendroides
37	55.5	42.4	709	5	09XUJ7 dendroides
38	54.5	41.6	66	2	076959 dendroides
39	54	41.2	105	4	09BYP8 dendroides
40	54	41.2	807	16	09KF73 dendroides
41	53.5	40.8	230	11	064507 dendroides
42	53.5	40.8	350	5	094589 dendroides
43	53	40.5	390	17	027680 dendroides
44	53	40.5	1110	3	092198 dendroides
45	53	40.5	1232	5	096088 dendroides

ALIGNMENTS

RESULT 1					
ID	09U744	PRELIMINARY;	PRT;	124	AA.
AC	09U744:				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	THERMAL HYSTERESIS PROTEIN ISOFORM 5-15 PRECURSOR.				
OS	Tenebrio molitor (Yellow mealworm).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;				
OC	Cucujiformia; Tenebrionidae; Tenebrio.				
OX	NCBI_TaxID=7067;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=FATBODY;				
RX	MEDLINE=99400482; PubMed=10471292;				
RA	Liu Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;				
RT	"A complex family of highly heterogeneous and internally repetitive				
RT	hyperactive antifreeze proteins from the beetle Tenebrio molitor.";				
RL	Biochemistry 38:11415-11424(1999).				
DR	EMBL; AF159118; AAD55260.1; -.				
DR	InterPro: IPR003460; AFP.				
DR	Pfam; PF02420; AFP; 9.				
KW	Signal.				
FT	SIGNAL				
FT	CHAIN 1 28				
FT	SEQUENCE 124 AA; 13005 MW; 8EC6871047FE2890B CRC64;				

POTENTIAL. THERMAL HYSTERESIS PROTEIN ISOFORM 5-15.

Query Match 96.2%; Score 126; DB 5; Length 124;
Best Local Similarity 87.0%; Pred. No. 2.1e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGXADCTCTACTGCGXCPNA 24
DB 30 CTGXADCTCTACTGCGXCPNA 52

RESULT 2

016121
ID 016121 PRELIMINARY; PRT: 124 AA.
AC 016121;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-3 (3-8) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF160496; AAB70752.1; -
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 9.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 124 THERMAL HYSTERESIS PROTEIN ISOFORM YL-3
SQ SEQUENCE 124 AA; 12993 MW; 0DEP87105613791A CRC64;
Query Match 96.2%; Score 126; DB 5; Length 124;
Best Local Similarity 87.0%; Pred. No. 2.1e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGADCTCTACTGCGCPCNA 24
DB 30 CTGADCTCTACTGCGCPCNA 52
RESULT 3
016122
ID 016122 PRELIMINARY; PRT: 148 AA.
AC 016122;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-4 (2-20) PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=97429943; PubMed=9285581;
RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
RT "Hyperactive antifreeze protein from beetles.";
RL Nature 388:727-728(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF160497; AAB70753.1; -
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 11.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 148 THERMAL HYSTERESIS PROTEIN ISOFORM YL-4
FT CHAIN 148 148 (2-20).
SQ SEQUENCE 148 AA; 15311 MW; EF07B36B8684365 CRC64;
Query Match 96.2%; Score 126; DB 5; Length 148;
Best Local Similarity 87.0%; Pred. No. 2.5e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGADCTCTACTGCGCPCNA 24
DB 30 CTGADCTCTACTGCGCPCNA 52
RESULT 4
090748
ID 090748 PRELIMINARY; PRT: 112 AA.
AC 090748;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 4-9 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159114; AAD55256.1; -
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9.
SQ SEQUENCE 112 AA; 11532 MW; 8A6124C6D31D0E19 CRC64;
Query Match 95.4%; Score 125; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 2.7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGADCTCTACTGCGCPCNA 24
DB 30 CTGADCTCTACTGCGCPCNA 52
RESULT 5
090746
ID 090746 PRELIMINARY; PRT: 112 AA.
AC 090746;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 3-4 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;

DR EMBL; AF160497; AAB70753.1; -
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 11.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 148 THERMAL HYSTERESIS PROTEIN ISOFORM YL-4
FT CHAIN 148 148 (2-20).
SQ SEQUENCE 148 AA; 15311 MW; EF07B36B8684365 CRC64;
Query Match 96.2%; Score 126; DB 5; Length 148;
Best Local Similarity 87.0%; Pred. No. 2.5e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGADCTCTACTGCGCPCNA 24
DB 30 CTGADCTCTACTGCGCPCNA 52
RESULT 4
090748
ID 090748 PRELIMINARY; PRT: 112 AA.
AC 090748;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 4-9 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159114; AAD55256.1; -
DR InterPro: IPR003460; AFP.
DR Pfam: PF02420; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 4-9.
SQ SEQUENCE 112 AA; 11532 MW; 8A6124C6D31D0E19 CRC64;
Query Match 95.4%; Score 125; DB 5; Length 112;
Best Local Similarity 87.0%; Pred. No. 2.7e-10;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 CTGADCTCTACTGCGCPCNA 24
DB 30 CTGADCTCTACTGCGCPCNA 52
RESULT 5
090746
ID 090746 PRELIMINARY; PRT: 112 AA.
AC 090746;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM 3-4 PRECURSOR.
OS Tenebrio molitor (Yellow mealworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=FATBODY;
RX MEDLINE=99400482; PubMed=10471292;

RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
 RT "A complex family of highly heterogeneous and internally repetitive
 hyperactive antifreeze proteins from the beetle *Tenebrio molitor*.";
 RL Biochemistry 38:11415-11424(1999).
 DR EMBL: AF159116; AAD55258.1; -.
 DR InterPro: IPR003460; AFP.
 DR Pfam: PF02420; AFP; 8.
 KW Signal.
 FT CHAIN 1 28 POTENTIAL.
 FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM 3-4.
 SQ SEQUENCE 112 AA; 11637 MW; 58994CCDCFF0F805 CRC64;
 QY 2 CTGAGDCTGCTGACTGCGCPCNA 24
 DB 30 CTGAGDCTGCTGACTGCGCPCNA 52

RESULT 6
 ID 016119 PRELIMINARY; PRT; 112 AA.
 AC 016119;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-1 (2-14) PRECURSOR.
 OS *Tenebrio molitor* (Yellow mealworm).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Tenebrionidae; Tenebrio.
 OX NCBI_TaxID=7067;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FATBODY;
 RX MEDLINE=97429943; PubMed=9285581;
 RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
 RT "Hyperactive antifreeze protein from beetles.";
 RL Nature 388:727-728(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FATBODY;
 RX MEDLINE=99400482; PubMed=10471292;
 RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
 RT "A complex family of highly heterogeneous and internally repetitive
 hyperactive antifreeze proteins from the beetle *Tenebrio molitor*.";
 RL Biochemistry 38:11415-11424(1999).
 DR EMBL: AF160494; AAB70750.1; -.
 DR InterPro: IPR003460; AFP.
 DR Pfam: PF02420; AFP; 8.
 KW Signal.
 FT CHAIN 1 28 POTENTIAL.
 FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM YL-1
 SQ SEQUENCE 112 AA; 11558 MW; C06D7AB6401BC909 CRC64;
 QY 2 CTGAGDCTGCTGACTGCGCPCNA 24
 DB 30 CTGAGDCTGCTGACTGCGCPCNA 52

RESULT 7
 ID 016120 PRELIMINARY; PRT; 112 AA.
 AC 016120;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE THERMAL HYSTERESIS PROTEIN ISOFORM YL-2 (1-3) PRECURSOR.
 OS *Tenebrio molitor* (Yellow mealworm).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Tenebrionidae; Tenebrio.
 OX NCBI_TaxID=7067;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FATBODY;
 RX MEDLINE=97429943; PubMed=9285581;
 RA Graham L.A., Liou Y.C., Walker V.K., Davies P.L.;
 RT "Hyperactive antifreeze protein from beetles.";
 RL Nature 388:727-728(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FATBODY;
 RX MEDLINE=99400482; PubMed=10471292;
 RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
 RT "A complex family of highly heterogeneous and internally repetitive
 hyperactive antifreeze proteins from the beetle *Tenebrio molitor*.";
 RL Biochemistry 38:11415-11424(1999).
 DR EMBL: AF160495; AAB70751.1; -.
 DR InterPro: IPR003460; AFP.
 DR Pfam: PF02420; AFP; 8.
 KW Signal.
 FT CHAIN 1 28 POTENTIAL.
 FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM YL-2
 SQ SEQUENCE 112 AA; 11667 MW; 5883E6DCDF0F805 CRC64;
 QY 2 CTGAGDCTGCTGACTGCGCPCNA 24
 DB 30 CTGAGDCTGCTGACTGCGCPCNA 52

RESULT 8
 ID 090747 PRELIMINARY; PRT; 112 AA.
 AC 090747;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE THERMAL HYSTERESIS PROTEIN ISOFORM D-16 PRECURSOR.
 OS *Tenebrio molitor* (Yellow mealworm).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Tenebrionidae; Tenebrio.
 OX NCBI_TaxID=7067;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FATBODY;
 RX MEDLINE=99400482; PubMed=10471292;
 RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
 RT "A complex family of highly heterogeneous and internally repetitive
 hyperactive antifreeze proteins from the beetle *Tenebrio molitor*.";
 RL Biochemistry 38:11415-11424(1999).
 DR EMBL: AF159115; AAD55257.1; -.
 DR InterPro: IPR003460; AFP.
 DR Pfam: PF02420; AFP; 8.
 KW Signal.
 FT CHAIN 1 28 POTENTIAL.
 FT SIGNAL 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM D-16.
 SQ SEQUENCE 112 AA; 11591 MW; BA290B5DB8B06C67 CRC64;
 QY 2 CTGAGDCTGCTGACTGCGCPCNA 24
 DB 30 CTGAGDCTGCTGACTGCGCPCNA 52

Query Match 93.1%; Score 122; DB 5; Length 112;
 Best Local Similarity 82.6%; Pred. No. 6.9e-10;
 Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXCPNA 24
111 :111111 1111 1111
DB 30 CTGSDCTCTACTGCGXCPNA 52

RESULT 9

Q90745 PRELIMINARY; PRT; 112 AA.
AC Q90745;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THERMAL HYSTERESIS PROTEIN ISOFORM C-9 PRECURSOR.
OS Tenebrio molitor (yellow mealworm)
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Tenebrionidae; Tenebrio.
OX NCBI_TaxID=7067;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FAIBODY;
RX MDLINE=99400482; PubMed=10471292;
RA Liou Y.C., Thibault P., Walker V.K., Davies P.L., Graham L.A.;
RT "A complex family of highly heterogeneous and internally repetitive
hyperactive antifreeze proteins from the beetle Tenebrio molitor.";
RL Biochemistry 38:11415-11424(1999).
DR EMBL; AF159117; AAD5259.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
KW Signal.
FT CHAIN 1 28 POTENTIAL.
FT SEQUENCE 29 112 THERMAL HYSTERESIS PROTEIN ISOFORM C-9.
SQ SEQUENCE 112 AA; 12002 MW; A4CA9DFB0AF81FB CRC64;

Query Match 84.0%; Score 110; DB 5; Length 112;
Best Local Similarity 78.3%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXCPNA 24
111 :111111 111 1111
DB 30 CTGADCTCTACTGCGXCPNA 52

RESULT 10

Q9NC09 PRELIMINARY; PRT; 123 AA.
AC Q9NC09;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ANTIFREEZE PROTEIN 10 (FRAGMENT).
GN AFP-10.
OS Dendroides canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroides.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroides canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179414; AAF86365.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
FT NON_TER 1 1
SQ SEQUENCE 123 AA; 12882 MW; A5D92CEAD81B4DA5 CRC64;

Query Match 78.6%; Score 103; DB 5; Length 123;
Best Local Similarity 69.6%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXCPNA 24
111 :111111 111 1111
DB 21 CTGSDCTCTACTGCGXCPNA 43

RESULT 11

Q9NC09 PRELIMINARY; PRT; 129 AA.
AC Q9NC09;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ANTIFREEZE PROTEIN 9.
GN AFP-9.
OS Dendroides canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroides.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroides canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179413; AAF86364.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 8.
DR Signal; 129 AA; 13679 MW; 5BE91630D6E8F131 CRC64;
SQ SEQUENCE 129 AA; 13679 MW; 5BE91630D6E8F131 CRC64;

Query Match 77.9%; Score 102; DB 5; Length 129;
Best Local Similarity 69.6%; Pred. No. 4.2e-07;
Matches 16; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXCPNA 24
111 :111111 111 1111
DB 27 CTGSDCTCTACTGCGXCPNA 49

RESULT 12

Q9NC08 PRELIMINARY; PRT; 148 AA.
AC Q9NC08;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ANTIFREEZE PROTEIN 11.
GN AFP-11.
OS Dendroides canadensis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Pyrochroidae; Dendroides.
OX NCBI_TaxID=55100;
RN [1]
RP SEQUENCE FROM N.A.
RA Andorfer C.A., Duman J.G.;
RT "Isolation and characterization of cDNA clones encoding antifreeze
RT proteins of the pyrochroid beetle Dendroides canadensis.";
RL J. Insect Physiol. 46:365-372(2000).
DR EMBL; AF179415; AAF86366.1; -.
DR InterPro; IPR003460; AFP.
DR Pfam; PF02420; AFP; 10.
SQ SEQUENCE 148 AA; 15290 MW; 3A01DEBD4D282A2CA CRC64;

Query Match 77.9%; Score 102; DB 5; Length 148;
Best Local Similarity 69.6%; Pred. No. 4.7e-07;
Matches 16; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 CTGADCTCTACTGCGXCPNA 24
111 :111111 111 1111
DB 27 CTGSDCTCTACTGCGXCPNA 49

RESULT 13
 Q9NCR1 PRELIMINARY; PRT; 121 AA.
 AC Q9NCR1: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ANTI-FREEZE PROTEIN 8.
 GN AFP-8
 OS Dendroidea canadensis.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Pyrochroidae; Dendroidea.
 OX NCBI_TaxID=55100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Andorfer C.A., Duman J.G.;
 RT "Isolation and characterization of cDNA clones encoding antifreeze
 RT proteins of the pyrochroid beetle Dendroidea canadensis."
 RL J. Insect Physiol. 46:365-372(2000).
 DR EMBL; AF179412; AAF86363.1; -.
 DR HSSP; P80719; ITXM.
 DR InterPro; IPR003460; AFP.
 DR Pfam; PF02420; AFP; 8.
 SQ SEQUENCE 121 AA; 12532 MW; D03516CF09C39004 CRC64;

Query Match 77.1%; Score 101; DB 5; Length 121;
 Best Local Similarity 69.6%; Pred. No. 5.5e-07;
 Matches 16; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 2 CTGXADCTGCTXACTGCGXCPNA 24
 DB 27 CTGGSDDCSCCTGACTTCNCNCPNA 49

RESULT 14
 Q9NCR3 PRELIMINARY; PRT; 91 AA.
 AC Q9NCR3: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ANTI-FREEZE PROTEIN 6 (FRAGMENT).
 GN AFP-6.
 OS Dendroidea canadensis.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Pyrochroidae; Dendroidea.
 OX NCBI_TaxID=55100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Andorfer C.A., Duman J.G.;
 RT "Isolation and characterization of cDNA clones encoding antifreeze
 RT proteins of the pyrochroid beetle Dendroidea canadensis."
 RL J. Insect Physiol. 46:365-372(2000).
 DR EMBL; AF179410; AAF86361.1; -.
 DR InterPro; IPR003460; AFP.
 DR Pfam; PF02420; AFP; 6.
 FT NON_TER 1
 SQ SEQUENCE 91 AA; 9536 MW; 5B91DEF5283FBA16 CRC64;

Query Match 76.3%; Score 100; DB 5; Length 91;
 Best Local Similarity 65.2%; Pred. No. 6.1e-07;
 Matches 15; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 2 CTGXADCTGCTXACTGCGXCPNA 24
 DB 22 CTGGSDDCSCCTGCTCNCNCPNA 44

AC Q9NCR5; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ANTI-FREEZE PROTEIN 3.
 GN AFP-3.
 OS Dendroidea canadensis.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Pyrochroidae; Dendroidea.
 OX NCBI_TaxID=55100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Andorfer C.A., Duman J.G.;
 RT "Isolation and characterization of cDNA clones encoding antifreeze
 RT proteins of the pyrochroid beetle Dendroidea canadensis."
 RL J. Insect Physiol. 46:365-372(2000).
 DR EMBL; AF179408; AAF86359.1; -.
 DR InterPro; IPR003460; AFP.
 DR Pfam; PF02420; AFP; 6.
 SQ SEQUENCE 104 AA; 11117 MW; A98AA2A184A0821A CRC64;

Query Match 74.8%; Score 98; DB 5; Length 104;
 Best Local Similarity 65.2%; Pred. No. 1.3e-06;
 Matches 15; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 2 CTGXADCTGCTXACTGCGXCPNA 24
 DB 26 CTGAPDCSACTTACTCVCNCPNA 48

Search completed: October 21, 2002, 16:39:36
 Job time : 9.76471 secs

